



Grado en ingeniería en tecnologías industriales: Organización industrial

Trabajo fin de grado:

**Gráficos de control basados en la
variación de un rango multivariante**

Autora: Carolina Canales Florencio

Tutor: Ignacio Cascos Fernández

Año: 2016

*A todos y cada uno de los que han confiado en mí
y en especial a mi familia por su apoyo e incondicionalidad*

Resumen

El presente TFG consiste en implementar una serie de gráficos de control para observaciones multivariantes a partir de un rango basado en la centralidad de los datos.

Este tipo de gráficos de control se inspiran en los que introdujo Liu [8], en los cuales para cada observación multivariante se obtiene su rango, es decir la proporción de observaciones históricas que son menos profundas que ella.

Para una muestra, de carácter multivariante, y un conjunto de observaciones históricas se obtienen los rangos según la centralidad de las observaciones de la muestra respecto de las históricas. Si las observaciones de la muestra son muy centrales respecto de las históricas sus rangos serán altos, mientras que si son mayoritariamente periféricas bajos. Estos rangos se pueden transformar de varias maneras para resumir su información, a lo largo de este trabajo se sumarán, multiplicarán y sumarán sus transformaciones Chi-cuadrado.

La implementación de los gráficos de control que se han descrito se ha realizado con el programa estadístico R, de libre distribución.

Con el desarrollo de este proyecto se puede ver cuál de los gráficos detallados proporcionan mejores resultados y las diferencias que aparecen entre ellos.

Índice de contenidos

Página

Capítulo I: Introducción, objetivos y estructura del proyecto.....	1
1.1 Introducción y objetivos.....	2
1.2 Estructura del documento.....	3
Capítulo II: Conceptos básicos en el control de procesos.....	4
2.1 Control estadístico de procesos.....	5
2.2 Gráficos de control.....	6
2.3 Gráficos de control de Shewhart.....	7
2.3.1 Gráficos de control para \bar{x} y R.....	7
2.3.2 Gráficos de control para \bar{x} y S.....	9
2.4 Gráficos CUSUM.....	10
2.5 T^2 Hotelling.....	10
2.5.1 Observaciones individuales.....	10
2.5.2 Medias muestrales.....	11
Capítulo III: Observaciones multivariantes.....	12
3.1 Procesos multivariantes.....	13
3.2 Gráficos descritos por Liu.....	14
3.2.1 Profundidad.....	14
3.2.2 Rango.....	17
Capítulo IV: Gráficos para rangos multivariantes.....	18
4.1 Gráficos Q.....	19
4.1.1 Promedio de rangos.....	20
4.1.2 Multiplicación de rangos.....	20
4.1.3 Conversión de rangos en Chi-cuadrado.....	21
4.1.4 Diferencias entre gráficos.....	22
4.2 Gráficos S.....	23
4.2.1 Promedio de rangos.....	23
4.2.2 Multiplicación de rangos.....	24
4.2.3 Conversión de rangos en Chi-cuadrado.....	25
4.2.4 Diferencias entre gráficos.....	26
Capítulo V: Ejemplos.....	27
5.1 Descripción de las simulaciones.....	28
5.2 Comentarios de los gráficos.....	35
Capítulo VI: Comparativa.....	36
6.1 Uso de experimentos.....	37
6.2 Descripción de las simulaciones.....	37
6.2.1 Variación de la media.....	38
6.2.2 Variación de la desviación estándar.....	39
Capítulo VII: Conclusiones.....	42
7.1 Conclusiones.....	43

Bibliografía.....	44
Anexos.....	45
Anexo 1: Detalle de las funciones.....	46
Anexo 2: Códigos en R utilizados.....	48
Anexo 3: Tabla de constantes para gráfico de control.....	54
Anexo 4: Valores de los rangos de los ejemplos de Liu.....	55

Índice de figuras

Página

Figura 1: Región de control con límites independientes para cada variable	
Control estadístico de la calidad - Montgomery.....	13
Figura 2: Gráfica de calor profundidad espacial.....	15
Figura 3: Gráfica de calor profundidad espacial correlación positiva.....	16
Figura 4: Gráfico de control Q promedio $N(1, 1)$	20
Figura 5: Gráfico de control Q producto $N(1, 1)$	21
Figura 6: Gráfico de control Q Chi-cuadrado $N(1, 1)$	22
Figura 7: Gráfico de control S promedio $N(1, 1)$	24
Figura 8: Gráfico de control S producto $N(1, 1)$	25
Figura 9: Gráfico de control S Chi-cuadrado $N(1, 1)$	26
Figura 10: Gráfico de control Q promedio rangos Liu.....	29
Figura 11: Gráfico de control Q producto rangos Liu.....	29
Figura 12: Gráfico de control Q Chi-cuadrado rangos Liu.....	30
Figura 13: Gráfico de control S promedio rangos Liu.....	30
Figura 14: Gráfico de control S producto rangos Liu.....	31
Figura 15: Gráfico de control S Chi-cuadrado rangos Liu.....	31
Figura 16: Gráfico de control Q promedio $N(2, 4)$	32
Figura 17: Gráfico de control Q producto $N(2, 4)$	32
Figura 18: Gráfico de control Q Chi-cuadrado $N(2, 4)$	33
Figura 19: Gráfico de control S promedio $N(2, 4)$	33
Figura 20: Gráfico de control S producto $N(2, 4)$	34
Figura 21: Gráfico de control S Chi-cuadrado $N(2, 4)$	34
Figura 22: Tabla comparativa variación en la media (1000 datos históricos).....	38
Figura 23: Tabla comparativa variación en la media (400 datos históricos).....	39
Figura 24: Tabla comparativa variación en la desviación (1000 datos históricos)....	40
Figura 25: Tabla comparativa variación en la desviación (400 datos históricos)....	40

CAPÍTULO I

INTRODUCCIÓN, OBJETIVOS Y ESTRUSCTURA DEL PROYECTO

1.1 INTRODUCCIÓN Y OBJETIVOS

El desarrollo de este proyecto se basa en la implementación de un software para la representación de los gráficos de control basados en la variación de un rango multivariante mediante el programa estadístico R ([3], [6]).

El programa R, con el que se han realizado todos los gráficos que aparecen en el estudio, es de código abierto enfocado básicamente a la estadística y el cual ofrece una gran cantidad de herramientas según sea necesario para cada usuario.

En este caso los paquetes estadísticos que se han utilizado son el *qcc* y el *ddalpha*.

El paquete *qcc* proporciona diversidad de gráficos de control y herramientas para el seguimiento y control de procesos, además de ofrecer la posibilidad de crear unos gráficos de control propios. Respecto al paquete *ddalpha* dispone de gran diversidad de métodos de cálculo de las profundidades de un conjunto de datos multivariados.

Los procesos multivariantes son de los más presentes en la vida real y por ello de gran utilidad. Se usan cuando en una misma observación están presentes varias variables que comúnmente están ligadas y por tanto se desea conocer la influencia de cada una de ellas en un determinado proceso.

Como estudiar cada variable de manera individual no es lo correcto y puede llevar a errores, se hace uso de este tipo de gráficos. Los datos tienen que ser tratados y para ello se usa el concepto de profundidad. Como se verá en adelante hay varios tipos de profundidad que son usadas según la finalidad que se vaya a dar al estudio. En nuestro caso se ha usado la profundidad simplicial.

La profundidad sirve para clasificar los datos según su grado de similitud con otros de referencia. Además para conocer la bondad de cada dato se hace uso de los rangos, ya que estos proporcionan la cantidad de muestras que son menores que la que se está estudiando.

En este proyecto se exponen una serie de gráficos que están basados en los gráficos Q y S, del tipo *X-bar* y CUSUM respectivamente, que propuso Liu en [8]. En el proyecto se usan estos gráficos y no los clásicos porque pueden ser usados para cualquier distribución, es decir están basados en la estadística no paramétrica, en cambio los gráficos clásicos tienen que ser usados con distribuciones normales.

Mediante estos gráficos se pretende conocer la bondad de cada uno de ellos según la transformación de los rangos que se haga. Así como conocer cuál de ellos proporciona mejores o peores resultados según las situaciones en las que se estén trabajando.

Los tratamientos de los rangos que se exponen en este estudio son promediarlos, multiplicarlos y transformarlos en una Chi-cuadrado y las diferentes circunstancias que se contemplan en este estudio son; la diferencia entre las distribuciones que se están comparando tanto en la media como en la desviación, la cantidad de datos iniciales que se tienen y la proporción de muestras erróneas aun sin que exista diferencia alguna entre las distribuciones de los datos que se comparan.

1.2 ESTRUCTURA DEL DOCUMENTO

A continuación se expone la forma en la que está estructurado el documento y un breve resumen de los contenidos tratados.

Capítulo II:

En este apartado se expone una introducción al control de procesos y se detallan los gráficos de control clásicos más relevantes, tanto en el caso univariante como en el multivariante.

Capítulo III:

En este capítulo se desarrolla el concepto de observaciones multivariantes así como el de profundidad y rango. Además se exponen algunas de las profundidades que hay y se detalla la que se va a seguir en el proyecto.

Capítulo IV:

En esta sección se detallan los gráficos multivariantes tipo Q y tipo S y los distintos tratamientos que se les dan a los rangos para estudiar su bondad, así como su representación.

Capítulo V:

En este capítulo se describen los experimentos que Liu realizó y se implementan siguiendo los mismos datos de entrada.

Capítulo VI:

En este apartado se exponen los distintos resultados que se han obtenido durante el estudio. Se realizan simulaciones con distintos parámetros en las distribuciones de entrada para poder ver el comportamiento de cada uno de los gráficos.

Capítulo VII:

En este capítulo se exponen las distintas conclusiones a las que se han llegado durante el estudio.

CAPÍTULO II

CONCEPTOS BÁSICOS EN EL CONTROL DE PROCESOS

2.1 CONTROL ESTADÍSTICO DE PROCESOS

El control estadístico de procesos (SPC) ([1], [2]) es un conjunto de herramientas que ayudan a buscar la estabilidad e intentar reducir la variabilidad de los procesos, desarrollando así la mejora continua y la productividad. Sus principales herramientas son:

- El histograma
- La hoja de verificación
- La gráfica de Pareto
- El diagrama de causa y efecto
- El diagrama de concentración de defectos
- El diagrama de dispersión
- El gráfico de control

De entre estas, aunque en este estudio se van a tratar más detalladamente los gráficos de control, hay que destacar el histograma y el diagrama de dispersión.

El histograma es comúnmente usado para la representación de variables univariantes continuas. Esta herramienta distribuye mediante rectángulos la frecuencia con la que aparece un determinado valor o un conjunto de valores, aunados mediante intervalos, con lo que proporciona de una manera muy clara e intuitiva los valores que son más propensos a darse.

En cuanto al diagrama de dispersión, es un instrumento que representa mediante coordenadas cartesianas el valor de dos variables, una en cada eje. Con él se puede observar si hay cierta relación entre ambas.

En el entorno en el que se está explicando, es utilizado a la hora de la representación de las variables multivariantes.

Los procesos están sometidos a numerosos factores, que son los que les dan esa variabilidad que se busca reducir mediante el SPC.

Como norma general se pueden diferenciar las causas de las variaciones en dos tipos:

No asignables- las cuales hacen referencia a todas las variaciones que están presentes en un proceso y que pueden tener ciertas oscilaciones, pero que no sobresale ninguna frente a las demás. Tienen un carácter aleatorio.

Asignables- en este tipo, de manera contraria al anterior, se pueden clasificar como tal a las causas que tienen un efecto preponderante frente a las demás. Estas son las que interesa descubrir y eliminar para mejorar el proceso.

Otro concepto clave es el hecho de que un proceso se encuentre bajo control. Este término indica que el proceso no sufre ninguna modificación causada por variables asignables y por lo tanto se puede asumir que el proceso permanece estable y seguirá la distribución establecida, consecuentemente se pueden realizar predicciones.

De manera contraria, el proceso estará fuera de control cuando se ve representada alguna causa assignable o se observa que se está siguiendo algún patrón.

Como se puede intuir, otro objetivo importante al que tiene que hacer frente el SPC es detectar con rapidez cuando aparecen las causas asignables en un proceso, para actuar de manera inminente.

En el estudio realizado, como ya se ha señalado, se estudia la principal de estas herramientas, los gráficos de control. Los cuales son una representación de cada uno de los datos que se manejan y de los valores límites que podrían tomar para que el proceso permanezca controlado.

Estos, mediante la representación de las observaciones, permiten mantener el proceso bajo control, reduciendo la variabilidad, y alertarnos de las variaciones que se puedan producir y en que medida afectan los diferentes factores al proceso, para estudiar como evitarlos y corregirlos.

No es posible hacer referencia al SPC y los gráficos de control y no detenerse a hablar de Walter Andrew Shewhart, el precursor de estas herramientas. El cual en tiempos donde lo único que se controlaba eran los productos, mediante su inspección, tuvo la brillantez de crear los gráficos de control como herramienta para intentar clasificar las causas por las que se producían las variaciones en no asignables o asignables.

2.2 GRÁFICOS DE CONTROL

Como se ha indicado anteriormente, el gráfico de control es una herramienta que sirve para observar gráficamente los procesos y detectar las posibles anomalías rápidamente.

En este gráfico se representa mediante puntos, normalmente unidos para dar una visualización más clara de evolución en el tiempo, el valor de la medida o característica que se quiera estudiar para cada una de las muestras, siempre de manera sucesiva y con el mismo tamaño muestral.

La importancia de este gráfico además de por todas sus características, también recae en la facilidad a la hora de interpretarlos, ya que son muy intuitivos.

Otras de sus utilidades características son la estimación de parámetros del proceso como pueden ser la media o la desviación típica, lo cual proporciona la capacidad de este para cumplir con las especificaciones.

También evitan el ajustar los procesos innecesariamente, ya que conocen cuando la variabilidad que se observa es ruido, es decir proveniente de causas no asignables, o es una causa asignable y por lo tanto proporcionan información para el diagnóstico.

Además, suponen un progreso en el hecho de que se pueden conocer las variaciones que se han producido en el momento y por lo tanto la identificación de sus causas se realiza más fácilmente y se puede solucionar con mayor rapidez, mejorando la eficiencia y productividad al disminuir los desperdicios y el retrabajo.

Los gráficos de control como se ha detallado, son una herramienta de las más importantes y más utilizada sobretodo en el sector industrial, ya que su utilización supone una mejora y un ahorro a bastantes niveles.

Además también se grafican los límites de control superior e inferior y la línea central, este último concepto representa los supuestos valores que tendrían que tomar las muestras si no existiese variabilidad.

En cuanto a los límites de control, se puede decir que son los que van a marcar la frontera para clasificar el proceso como controlado o no. Aunque aquí hay que hacer un inciso, cabe señalar que si las muestras que se están representando tienen un claro signo de tendencia o siguen algún patrón, el proceso también se considera como fuera de control, ya que para que se considere un proceso como controlado, todas las muestras tienen que tener un patrón aleatorio. De la misma manera, podemos decir que un proceso está controlado aunque aparezca alguna muestra fuera de los límites establecidos, estas muestras son falsas alarmas y son debidas al carácter aleatorio de nuestras observaciones, ya que se puede generar alguna que sobrepase los límites pero de manera puntual.

2.3 GRÁFICOS DE CONTROL DE SHEWHART

Como se ha comentado, el precursor de esta herramienta es el físico, ingeniero y estadístico W. A. Shewhart. Él, a grandes rasgos, realizó este tipo de controles en el que estudiaba básicamente la media del proceso y su rango o desviación según fuera conveniente. Los procesos para los que sirven estos gráficos clásicos son los univariantes, es decir en los cuales únicamente se está midiendo y observando una variable.

Debido a los avances y estudios posteriores, se tuvo en cuenta la idea de que en una misma muestra mayoritariamente aparecen varias variables al mismo tiempo, las cuales normalmente están relacionadas, a este tipo de procesos se les denomina multivariantes y como consecuencia surgieron nuevos gráficos de control, en los cuales se centra este proyecto.

2.3.1 GRÁFICOS DE CONTROL PARA \bar{x} Y R

Se denominan como \bar{X} , son gráficos en los que se representa el progreso del promedio de las muestras, estas muestras normalmente son de tamaño n .

A su vez R hace referencia al rango de las observaciones, es decir a la diferencia entre el mayor y el menor valor de cada conjunto de muestras n y por lo tanto se denota como \bar{R} al promedio de dichos rangos.

En este tipo de gráficos se tiende a asumir la normalidad de las variables, por lo que su media también se distribuirá del mismo modo. Cuando la variable a estudiar no sigue una distribución normal y el tamaño de la muestra es grande, se puede asumir dicha normalidad mediante el teorema central del límite.

Los límites de control que se establecen típicamente son los siguientes:

$$LCS = \mu + 3 \frac{\sigma}{\sqrt{n}}$$

$$central = \mu$$

$$LCI = \mu - 3 \frac{\sigma}{\sqrt{n}}$$

Shewhart estableció los límites tres sigma y es el valor que comúnmente suelen tomar, esto se corresponde, cuando las observaciones son normales, con un valor del parámetro α de 0.0027. Se denomina α al nivel de significación, esto es la probabilidad de obtener falsas alarmas, es decir resultados anómalos aunque los datos tengan los mismos parámetros característicos que con los que se están midiendo. Además este parámetro puede ser modificado según el nivel de control que se desee tener, cuanto mayor sea se van a dar mayor cantidad de falsas alarmas, por lo tanto será un control mucho más basto.

Los valores tanto de la media como de la desviación estándar no suelen ser conocidos, por lo que se suelen estimar.

Las estimaciones se pueden realizar, y por lo tanto van a ser consideradas como buenas, cuando se tiene una cantidad de muestras considerables así como el tamaño muestral, para el primero tiene que estar como mínimo entorno a la veintena y para el segundo unas 4 ó 6 son suficientes.

El estimador de la media es:

$$\hat{\mu} = \bar{\bar{x}} = \frac{\sum_{i=1}^m \bar{x}_i}{m}$$

Siendo m la cantidad de muestras históricas y \bar{x}_i el promedio de la i -ésima muestra.

Mientras que para estimar la desviación estándar, hay que hacer uso de una serie de coeficientes y de los rangos de las muestras.

Dichos rangos son la diferencia entre el mayor y el menor valor de cada conjunto de muestras n .

Así la desviación estándar se puede aproximar a partir del promedio de los rangos y en cuanto a los coeficientes, vienen determinados en una tabla, la cual se puede consultar en el anexo 3 de este documento, para 3 desviaciones típicas del estimador del parámetro.

Finalmente los límites se caracterizan como:

$$LCS = \bar{\bar{x}} + A_2 \bar{R}$$

$$central = \bar{\bar{x}}$$

$$LCI = \bar{\bar{x}} - A_2 \bar{R}$$

Para el gráfico del rango los límites que se han establecido son los siguientes:

$$LCS = D_4 \bar{R}$$

$$central = \bar{R}$$

$$LCI = D_3 \bar{R}$$

De igual modo estos coeficientes también están tabulados y se encuentran en la tabla citada anteriormente.

El gráfico de los rangos, lo que representa es la evolución del rango propiamente dicho, es decir se puede ver la variabilidad del proceso.

Estos gráficos se complementan, ambos son necesarios para poder determinar un proceso como controlado.

Primeramente se representa el gráfico de los rangos y se estudia si está controlado y cuando lo está, se procede a desarrollar el de la media. Si ha sido necesario estimar los parámetros, hay que comprobar que no queda ninguna muestra fuera de los límites de control, si esto sucede se deben eliminar dichas muestras y variar los parámetros hasta que las muestras queden dentro de los límites, así el proceso estará controlado.

2.3.2 GRÁFICOS DE CONTROL PARA \bar{x} Y S

Se utiliza del mismo modo que el gráfico explicado anteriormente, pero ahora en vez de utilizar el concepto de rango para la estimación de la varianza, se hace con la desviación estándar.

Este gráfico se suele usar en menor medida que el anterior aunque estadísticamente sea más completo por el hecho de que trabaja con todas las muestras que se tienen y no únicamente con los valores extremos como ocurre en el otro tipo.

Por tanto se tiende a usar cuando el tamaño muestral es considerable, comúnmente mayor que 10.

De igual manera que en el gráfico anterior, se tiene dos gráficos para el estudio, el X-bar que representa la variación de la media de las muestras y el gráfico S, que va a representar la variabilidad del proceso.

Los límites de control, igual que en el gráfico anterior, también vienen regidos por varios coeficientes tabulados que se pueden consultar en la tabla del anexo 3. Los límites para la media son los siguientes:

$$LCS = \bar{\bar{x}} + A_3\bar{S}$$

$$central = \bar{\bar{x}}$$

$$LCI = \bar{\bar{x}} - A_3\bar{S}$$

Así para la desviación estándar vienen determinados por:

$$LCS = B_4\bar{S}$$

$$central = \bar{S}$$

$$LCI = B_3\bar{S}$$

S es la forma de denominar a la desviación estándar de la muestra y se calcula mediante la siguiente fórmula:

$$S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

Por lo tanto el símbolo \bar{S} hace referencia a la media de las desviaciones estándar.

2.4 GRÁFICOS CUSUM

Estos gráficos son una buena alternativa a los de Shewhart, debido a que incorporan la información de todas las muestras anteriores y no únicamente de la última como las de Shewhart hacen. Por lo tanto, permiten observar de una manera mucho más rápida si hay algún tipo de variación en las muestras aunque estas sean muy pequeñas, funcionan mejor que las de Shewhart cuando la desviación respecto a la media es menor que dos desviaciones típicas, por lo que son mucho más efectivas.

El gráfico consiste en la representación de las sumas acumuladas de la variación que tiene cada muestra o conjunto de estas respecto a un valor objetivo. Cuanto menor sea el tamaño de la muestra, el gráfico presentará una mayor eficacia.

Cuando el proceso está controlado los puntos deben fluctuar entorno al valor 0, es decir son observaciones semejantes al valor objetivo establecido. Así un proceso queda determinado como fuera de control cuando se observa que presenta cierta tendencia.

Para la representación se suele utilizar el *CUSUM tabular*, este gráfico va acumulando las desviaciones que se producen respecto a la media, pero las positivas y las negativas por separado, y siempre que la desviación sea mayor que un determinado valor umbral. Los límites de control se suelen establecer en cinco veces la desviación estándar del proceso.

La otra forma de representarlo es con la *máscara V*, los puntos se representan de la misma manera, mediante la acumulación de las sumas, pero en este caso se establecen los límites de control mediante dicha plantilla V y en el momento que las muestras interseccionan con dicha plantilla son consideradas fuera de control.

2.5 T² HOTELLING

Los gráficos T^2 de Hotelling [2] son una analogía de los gráficos de control de la media de Shewhart pero para variables múltiples.

Un problema principal de este gráfico es la falta de información que se tiene sobre la causa de la salida de un punto de los límites de control, por lo que se suele llevar en paralelo gráficos univariantes para cada una de las variables.

Son gráficos en los que las variables de estudio deben suponerse normalizadas, además según la distribución de los datos que manejemos, tenemos dos variaciones de estos gráficos, una si trabajamos con datos subagrupados u otra si nuestras observaciones son individuales.

2.5.1 OBSERVACIONES INDIVIDUALES

Las aproximaciones que se han realizado para este caso especial donde el tamaño de la muestra es 1, son de gran importancia debido al uso generalizado que se tiende a hacer en las compañías industriales de este tipo de variables.

Hay ciertas variaciones en cuanto a la estimación de los límites de control, en función de la cantidad de muestras (m) que sean objeto de estudio y de la cantidad de características (p) que se estén manipulando en cada una de las muestras.

Así el estadístico T^2 queda definido como:

$$T^2 = (x - \bar{x})'S^{-1}(x - \bar{x})$$

Los límites de control quedan determinados por:

$$LCS = \frac{p(m-1)}{m-p} F_{\alpha, p, m-p}$$

$$LCI = 0$$

Esta es la ecuación que se ha usado en este estudio, aunque existen otras posibilidades según las condiciones en las que se trabajen.

2.5.2 MEDIAS MUESTRALES

Se dan cuando el tamaño de la muestra es mayor que 1.

La ecuación que determina el estadístico T^2 es la siguiente:

$$T^2 = k(\bar{x} - \bar{\bar{x}})'S^{-1}(\bar{x} - \bar{\bar{x}})$$

Siendo \bar{x} el vector de la media muestral consecuentemente $\bar{\bar{x}}$ el vector de la media del proceso y S la matriz de varianzas-covarianzas de las observaciones.

A su vez, los límites de control vienen definidos por:

$$LCS = \frac{p(n-1)(k-1)}{nk - n - p + 1} F_{\alpha, p, nk-n-p+1}$$

$$LCI = 0$$

Siendo n el tamaño muestral, k la dimensión en la que se trabaja y F hace referencia a un cuantil de la distribución F la cual se corresponde con el cociente de dos varianzas, es decir compara la variabilidad entre los datos históricos y la nueva observación.

Estos gráficos están basados en el concepto de profundidad, el cual se verá en detalle en el siguiente capítulo.

CAPÍTULO III

OBSERVACIONES MULTIVARIANTES

3.1 PROCESOS MULTIVARIANTES

Como norma general se tiende al uso de tablas y gráficos que hacen referencia únicamente al análisis de una variable. Pero en la realidad este caso no es el que ocurre mayoritariamente, se tiende a estudiar procesos en los que se incluyen varias variables normalmente relacionadas. Este tipo de procesos, son los denominados multivariantes.

Generalmente la idea que se tiene para la gestión de este tipo de procesos es tomar cada una de las variables como independientes y utilizar los gráficos clásicos para el estudio individual de cada una de ellas.

A continuación se expone un ejemplo para aclarar el mal uso que se puede hacer al no utilizar los gráficos adecuados:

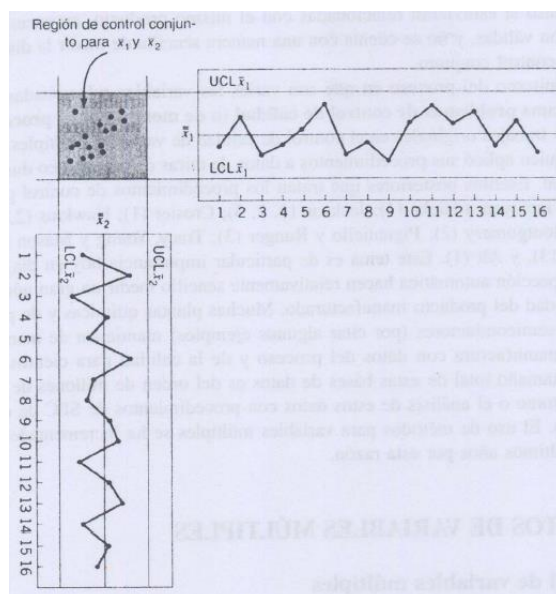


Figura 1: “Región de control con límites independientes para cada variable” Control estadístico de la calidad - Montgomery

Como se puede ver, si se estudia cada una de las variables independientemente, no se percibe ningún indicio de que alguna de las muestras esté fuera de control. Pero al representarlas conjuntamente aunque continúen dentro de los límites de control hay un punto que a priori parece un tanto inusual.

En este ejemplo se han usado los gráficos clásicos de Shewhart para la media y la probabilidad de que cada uno de los gráficos esté bajo control es de 0.9973, por lo tanto que esto ocurra de manera simultánea supone una probabilidad de 0.9973×0.9973 , o lo que es lo mismo 0.9946. Como se puede ver, una probabilidad bastante menor que la establecida.

De la misma manera ocurre con la probabilidad de que un punto sobrepase ambos límites de control, pasa de ser 0.0027 a 0.00000729.

Por lo que de este modo se puede ver la distorsión que puede aparecer al usar de manera inadecuada los gráficos según las variables que se estén tratando. Así cuanto mayor sea la cantidad de características que se estén midiendo, mayor será esta diferencia. La

probabilidad (α') de que un punto esté dentro de los límites viene determinada por la siguiente expresión:

$$1-\alpha'=(1-\alpha)^p$$

Siendo p la cantidad de características independientes.

Si las características que se miden no son independientes, la visualización no es tan evidente ya que no hay expresiones sencillas que lo muestren.

3.2 GRÁFICOS DESCRITOS POR LIU

Los gráficos que se van a llevar a cabo son los que fueron descritos por Liu [3] denominados como Q y S. Estos gráficos son utilizados para la monitorización de procesos multivariantes y se presentan como una alternativa a los propuestos por Hotelling.

El uso de gráficos multivariantes viene dado por los problemas que traen la utilización de otros tipos de gráficos como el \bar{X} , $\bar{\bar{X}}$ o el CUSUM al tratar determinadas observaciones que no cumplen con la asunción de normalidad o no son de carácter univariante, ya que en la realidad la mayoría de las observaciones que encontramos tienen varias variables para cada observación, son del tipo multivariante.

Estos gráficos están basados en el concepto de profundidad de los datos y en la ordenación mediante rangos para facilitar el estudio de cada una de las muestras y así identificar los procesos como controlados o no.

3.2.1 PROFUNDIDAD

La idea de profundidad ([4], [10]) se refiere al nivel de centralidad que tiene un nuevo dato en relación a una distribución establecida y por lo tanto permite ordenar nuestras muestras.

Por lo que para asignar dicha profundidad a cada nuevo dato, básicamente lo único necesario es conocer la distribución de la que proceden los datos con los que comparamos.

La función de profundidad es la que va a permitir asignar dicho grado de centralidad respecto a la distribución de referencia.

Dicha función debe cumplir las siguientes propiedades: estar acotada entre 0 y 1, ser invariante afín, es decir que la profundidad de un punto no dependa de las coordenadas del sistema elegido, anulación en el infinito, monotonidad y maximalidad al centro.

Hay varios tipos de profundidades, en este estudio se ha utilizado la espacial. Este tipo de profundidad se ha elegido para evitar los empates, que pueden surgir cuando los datos están en la periferia de la nube.

La profundidad espacial o L_1 otorga valores entre el 0 y el 1 a los nuevos datos, los cuales representan su centralidad respecto a la de referencia, es decir que ayudan a conocer su localización en la nube.

Se toman como datos más profundos a los que sus profundidades tienen valores más altos y por lo tanto serán los datos más centrales y acordes a la distribución, y poco profundos a los que sus valores son cercanos al 0 y consecuentemente muy dispares respecto a la nube de datos con la que se está comparando.

La fórmula de esta profundidad es:

$$D(x) = 1 - \left\| \sum_{i=1}^n \frac{x - y_i}{\|x - y_i\|} \right\|$$

Donde y_i hace referencia a una muestra de n observaciones históricas y x a cada nueva observación que se está tratando.

Con el propósito de ver de manera más clara el concepto de profundidad espacial, se presentan los siguientes gráficos:

La siguiente figura representa el gráfico de calor para la profundidad espacial para datos independientes con distribución $N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$.

En el centro toma el valor máximo y según se aleja, va disminuyendo.

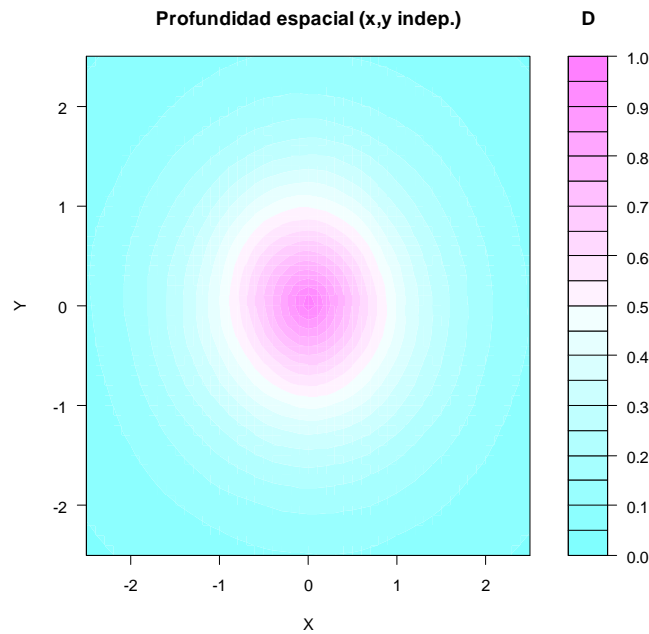


Figura 2: "Gráfica de calor profundidad espacial"

De igual manera se ha representado la profundidad de un conjunto de datos, pero ahora siguen la distribución $N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix}\right)$ para ver un ejemplo de correlación.

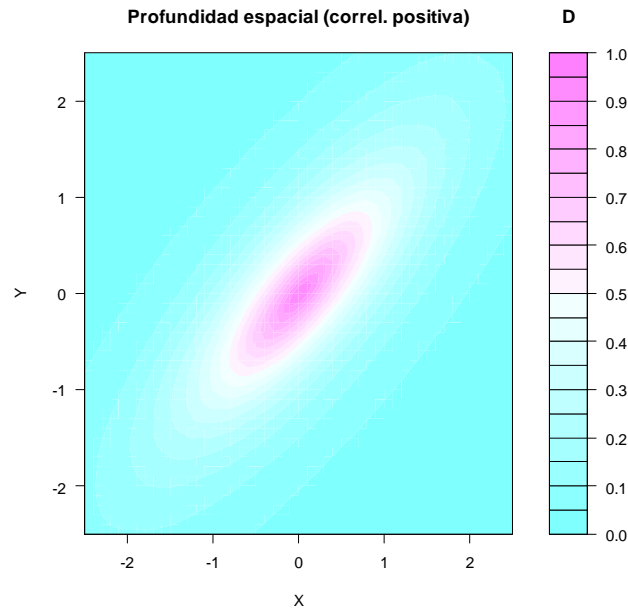


Figura 3: “Gráfica de calor profundidad espacial correlación positiva”

Existen más tipos de profundidad, algunas de las que destacan son la de Mahalanobis [7], la cual es una transformación de la distancia de Mahalanobis.

Otro concepto de profundidad importante es el de Tukey [5], la denominada profundidad semiespacial. Esta se concibe como la menor fracción de puntos de la muestra que hay en un semiespacio que contenga al dato que se está estudiando. A cada punto de estudio se le asigna como profundidad el mínimo de la probabilidad de los semiespacios que contienen al punto.

Un ejemplo para mejorar la comprensión de esta profundidad puede ser, si las muestras tienen carácter bidimensional se traza una recta que pase por uno de los puntos, creando dos conjuntos de muestras. La profundidad del dato estudiado será la probabilidad de que este pertenezca al conjunto donde hay menor número de muestras.

Y por último destacar la profundidad simplicial [9], la cual es usada por Liu en sus experimentos.

Esta se puede definir como la probabilidad de que un dato esté en el simplex cuyos vértices son $d+1$ observaciones independientes de una distribución, siendo d la dimensión en la que estamos trabajando.

Para tener una idea más clara de este concepto se recurre a un ejemplo. En el caso de que se estén tratando los datos en un espacio bidimensional, se forma un triángulo cuyos vértices pertenecen a las observaciones dadas y la profundidad será la probabilidad de que el nuevo dato esté dentro del triángulo construido.

3.2.2 RANGO

Tras fijar las profundidades a los nuevos datos, se clasifican en función de su centralidad. Para ello vamos a usar los rangos, los cuales dan una medida de la cantidad de datos que tienen profundidades iguales o menores a la de estudio. Por lo tanto en función de la profundidad que se use usando el valor de los rangos para cada observación varía. En este caso como es lógico también toman valores entre 0 y 1, donde los rangos altos representan datos periféricos y los bajos datos pertenecientes a la zona central, es decir, un valor pequeño para el rango indica que es un dato periférico y por lo tanto la similitud con la distribución de los históricos no es demasiada.

Para este caso particular la cantidad de datos históricos y nuevos que se han tomado son de 1000 y 200 respectivamente, ambos bidimensionales con una distribución normal, para la facilidad tanto de comprensión y visualización como de agilidad para la programación y el cómputo, pero puede ser modificable según las necesidades.

CAPÍTULO IV

GRÁFICOS PARA RANGOS MULTIVARIANTES

En el capítulo II se ha comentado el uso del gráfico T^2 de Hotteling para los procesos multivariantes. Como se ha dicho es uno de los más utilizados en este ámbito, pero también hay otras posibilidades. Este estudio se centra en el uso de los gráficos descritos por Liu [8], los cuales se detallan a continuación, para tratar este tipo de procesos.

En cuanto a las diferencias principales que presentan estos gráficos destacan que para los gráficos de Liu no es necesario que la distribución que siguen las variables sea una normal, por lo que su uso puede ser más generalizado. Aunque también hay que señalar que los gráficos de Hotteling son más eficientes cuando la desviación del nuevo dato es pequeña en comparación con los datos controlados de los que se disponen, pero cuando esta desviación es grande su eficiencia es prácticamente la misma.

4.1 GRÁFICOS Q

El gráfico Q descrito en el Liu es del tipo $X\text{-bar}$, los cuales son similares a los del caso univariante para \bar{X} . Estos grafican la media de cada muestra, a su vez compuesta por n observaciones.

Es decir, únicamente muestran la información relativa a la observación que se esté tratando en dicho momento.

Como no hacen ninguna referencia a la secuencia de observaciones anteriores, puede ocurrir que se den como buenas un conjunto de observaciones que muestren cierta tendencia. Para solventar este tipo de problemas, puede hacerse uso de límites de advertencia u otro tipo de reglas de sensibilidad, aunque bien es verdad que el uso de estas técnicas tiende a disminuir la sencillez característica de este tipo de gráficos, por lo que tampoco son del todo recomendables.

A la hora de proceder a la representación de estos gráficos, primeramente se asignan las profundidades a cada dato, en este caso utilizando la definición de profundidad espacial, y posteriormente se hallan los rangos.

En los casos de este estudio, en general se puede decir que el gráfico Q representa puntos formados por la manipulación de n observaciones de los rangos y tendrán unos límites de control constantes.

Para estudiar la anomalía de las muestras se combina la información que proporcionan los n rangos de diferentes maneras: promediando los rangos, multiplicando los rangos y convirtiéndolos en Chi-cuadrados.

En el caso de los ejemplos que se representan a continuación se ha decidido que los datos históricos sigan la distribución $N(0,1)$ y para la generación de las muestras anómalas se ha utilizado una $N(1,1)$ con el fin de que haya cierta variación apreciable pero tampoco una anomalía exagerada.

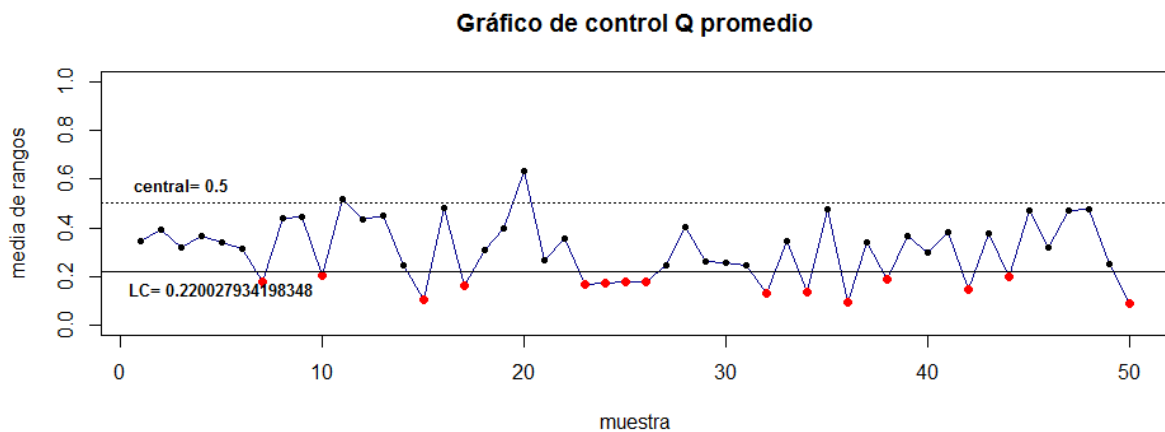
4.1.1 PROMEDIO DE RANGOS

Este es el caso que propone Liu, en él se hace la media de los rangos del conjunto de las n observaciones.

Las ecuaciones que caracterizan el límite control y el valor central varían en función del valor de n que se tome, en este estudio se está haciendo las simulaciones para $n=4$, por lo que la ecuación que los determina es la siguiente:

$$LC = \frac{(n! \alpha)^{1/n}}{n}$$

$$Central = 0.5$$



Comandos:

```
graf.controlQAv(new.rank,n=4,alpha=0.025)
rankd(new.depth,hist.depth)
depth.spatial(new.data, hist.data)
```

Figura 4: “Gráfico de control Q promedio $N(1, 1)$ ”

Como se puede ver, la muestra está fuera de control cuando es menor que el límite establecido, es decir si el promedio de los rangos es demasiado pequeño.

4.1.2 MULTIPLICACIÓN DE RANGOS

En este caso se puede decir que la muestra queda fuera de control cuando el producto de los n rangos (r_i) $\prod_{i=1}^n r_i$ es pequeño. Como al hacer este tipo de operaciones se trataría con valores poco manejables, se realiza un cambio y se transforman aplicando logaritmos.

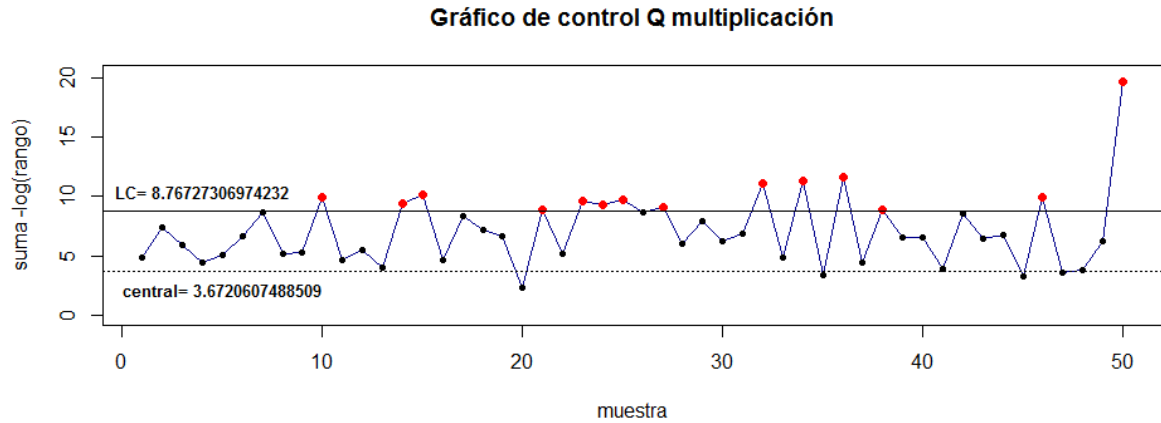
El rango r_i es uniforme en $(0, 1)$ y el $\log(1-r_i)$ sigue una distribución exponencial con parámetro $\lambda=1$. El logaritmo de los productos $\log(\prod_{i=1}^n (1-r_i))$ es la suma de los logaritmos $\sum_{i=1}^n \log(1-r_i)$ y sigue una distribución Gamma de parámetro n y 1 y rechazamos para valores grandes, es decir cuando el rango es pequeño.

El límite de control se establece en el cuantil $(1-\alpha)$ de la siguiente distribución:

$$LC = qgamma(1-\alpha, n, 1)$$

Y el valor central viene definido por:

$$central = qgamma(0.5, n, 1)$$



Comandos:

```
graf.controlQProd(new.rank,n=4,alpha=0.025)
rankd(new.depth,hist.depth)
depth.spatial(new.data, hist.data)
```

Figura 5: “Gráfico de control Q producto $N(1, 1)$ ”

4.1.3 CONVERSIÓN DE RANGOS EN CHI-CUADRADO

Para esta opción, se aplica el método de la transformada inversa para convertir una uniforme (0, 1) de (1- r) a una Chi-cuadrado con 1 grado de libertad, así sumando n de ellas, se tiene n grados de libertad.

La expresión que permite convertir los rangos en una Chi-cuadrado es $qchisq(1-r_i)$ y luego se suman los n valores correspondientes.

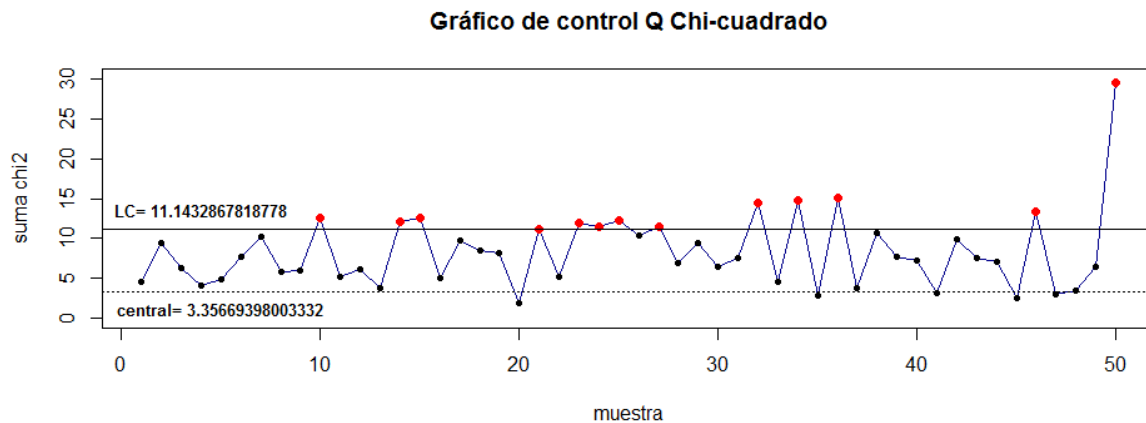
La anomalía es detectada cuando el valor de las muestras tras la conversión sea excesivamente grande, es decir se rechaza si el rango es pequeño.

El límite de control se establece en el cuantil (1- α) y viene determinado por la expresión:

$$LC = qchisq(1-\alpha, df=n)$$

Así como el valor central es:

$$Central = qchisq(0.5 df=n)$$



Comandos:
`graf.controlQChi2(new.rank,n=4,alpha=0.025)`
`rankd(new.depth,hist.depth)`
`depth.spatial(new.data, hist.data)`

Figura 6: “Gráfico de control Q Chi-cuadrado $N(1, 1)$ ”

4.1.4 DIFERENCIAS ENTRE GRÁFICOS

Los gráficos representan los datos generados con una cierta anomalía respecto a los que se han tomado como controlados. Estos datos anómalos siguen una distribución normal multivariante la cual varía en una unidad respecto a los datos con los que se están comparando.

Se estudian 50 muestras de tamaño 4, en cada uno de los gráficos anteriores los datos con los que se trabajan son los mismos y han sido ordenados y clasificados según los conceptos de profundidad y rango previamente explicados.

Como se puede ver en las gráficas, la cantidad de muestras que son detectadas como fuera de control son bastante similares en los tres procedimientos.

Para el gráfico del promedio se detectan 15 muestras como descontroladas, en el caso del producto la cantidad es de 14 y por último con la conversión en Chi-cuadrado de los datos se obtienen 13 muestras que sobrepasan el límite.

El que mayor cantidad de muestras anómalas detectadas verdaderamente como fuera de control es el caso en el que los datos se manipulan realizando el promedio de ellos, pero la diferencia como ya se ha dicho es minúscula. Aunque cabe decir que el número de muestras que se han generado hasta la detección de una primera muestra fuera de control es de 6, en comparación con las 9 que han sido necesarias para los otros procesos. Lo que se puede ver como un primer indicio del mal funcionamiento del proceso, es decir el promediar los rangos es más rápido que los otros procesos.

Las diferencias entre los gráficos del producto y de la Chi-cuadrado son bastante escasas, aunque se podría señalar un nivel más estricto en los límites de control del primero de estos.

El porcentaje de detección de anomalías en el proceso mediante el uso de estos gráficos está entorno al 28%, un dato verdaderamente bajo si se desea emplearlos en procesos para mejorar la eficacia.

4.2 GRÁFICOS S

Este gráfico es del tipo CUSUM es decir, para cada uno de los puntos que se quieren graficar, se va a utilizar toda la información disponible en ese momento.

Es un gráfico bastante bueno para detectar rápidamente cuando hay muestras anómalas con pequeñas variaciones, que se alejan de los valores centrales, pero que los gráficos Q no señalarían como anómalas al no traspasar el límite establecido.

Para la utilización de la información, se van a combinar los rangos de la misma manera que hemos hecho para los gráficos Q, es decir promediándolos, multiplicándolos y convirtiéndolos en Chi-cuadrado.

En estos gráficos, como la información va variando a lo largo del tiempo, los límites de control y el valor central también lo hacen.

Las distribuciones que se han elegido para los datos históricos y los anómalos son los mismos que para el gráfico anterior Q.

4.2.1 PROMEDIO DE RANGOS

La representación de cada uno de los puntos se lleva a cabo haciendo la media de todos los rangos anteriores que se tienen hasta ese momento.

En este caso se han señalado dos límites de control posibles según varíe el parámetro α , si α es menor o igual que $\frac{1}{i!}$ siendo i cada uno de los valores enteros desde 1 hasta la cantidad de nuevos rangos que se han generado, el valor del límite de control será el siguiente:

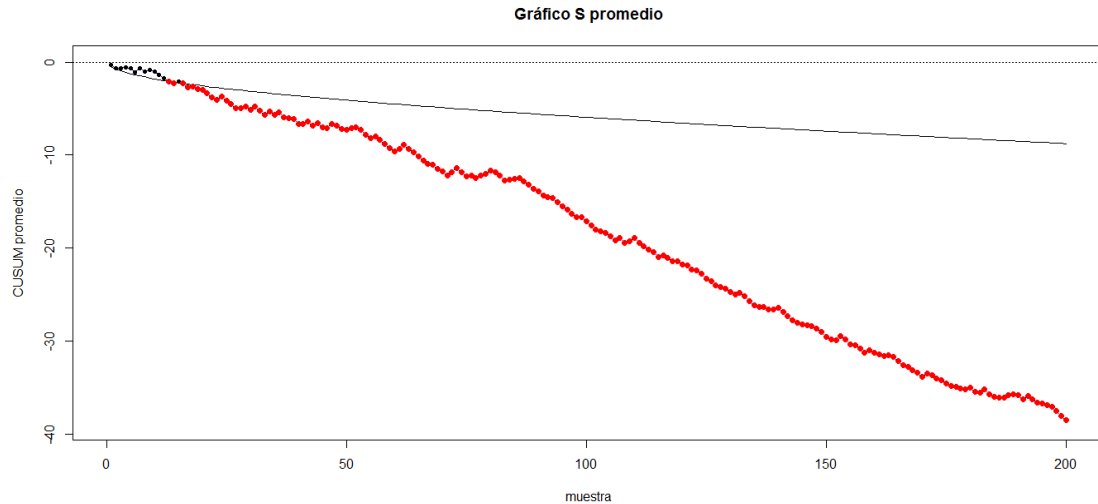
$$LC(i) = (i! \alpha)^{1/\alpha} - \frac{i}{2}$$

En el caso que α no tome dichos valores, el límite de control, el cual viene dado por la aproximación del teorema central del límite, es:

$$LC(i) = qnorm(\alpha) i \sqrt{\frac{\frac{1}{k} + \frac{1}{i}}{12}}$$

Siendo k la cantidad de datos iniciales que se tienen.

Para ambos casos el valor central es el mismo e igual a 0.



Comandos:

```
graf.controlSAv(new.rank,length(hist.depth),alpha=0.025)
rankd(new.depth,hist.depth)
depth.spatial(new.data, hist.data)
depth.spatial(hist.data, hist.data)
```

Fuera de control muestras 13, 14, 16...

Figura 7: “Gráfico de control S promedio $N(1, 1)$ ”

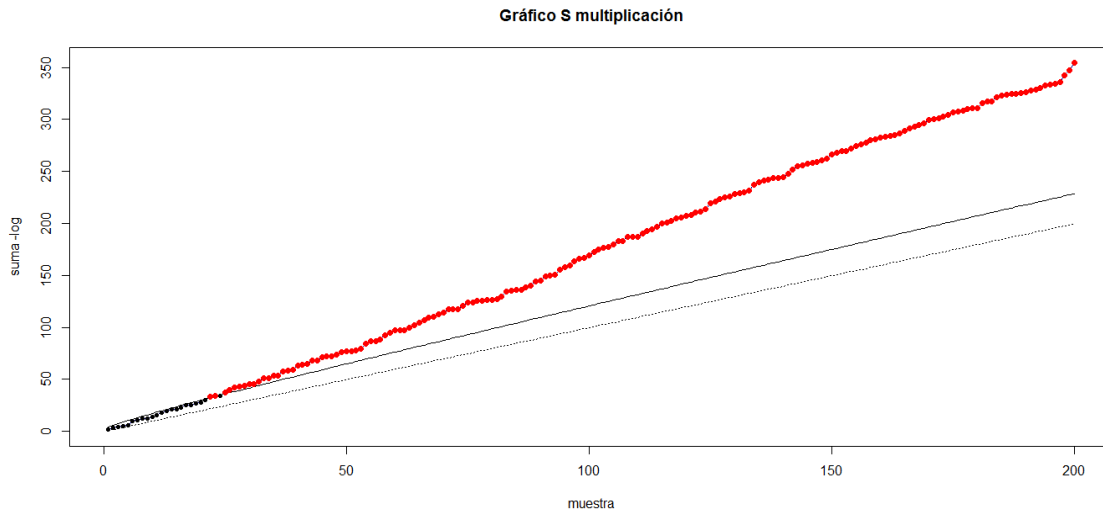
4.2.2 MULTIPLICACIÓN DE RANGOS

De igual modo que para el caso del gráfico Q, se va a utilizar el opuesto del logaritmo para la representación de las observaciones. Además aquí la combinación de los datos según avanza el tiempo se hace mediante el sumatorio de todos los valores de los que se disponen hasta ese instante.

Los valores del límite de control y el valor central también van variando y se recalculan para cada nueva observación según las ecuaciones:

$$LC(i) = qgamma(1 - \alpha, shape = i, rate = 1)$$

$$central(i) = qgamma(0.5, shape = i, rate = 1)$$



Comandos:
`graf.controlSProd(new.rank,alpha=0.025)`
`rankd(new.depth,hist.depth)`
`depth.spatial(new.data, hist.data)`

Fuera de control muestras 22, 23, 25...

Figura 8: “Gráfico de control S producto $N(1, 1)$ ”

4.2.3 CONVERSIÓN DE RANGOS EN CHI-CUADRADO

Este gráfico utiliza la siguiente función para realizar la conversión de los rangos:

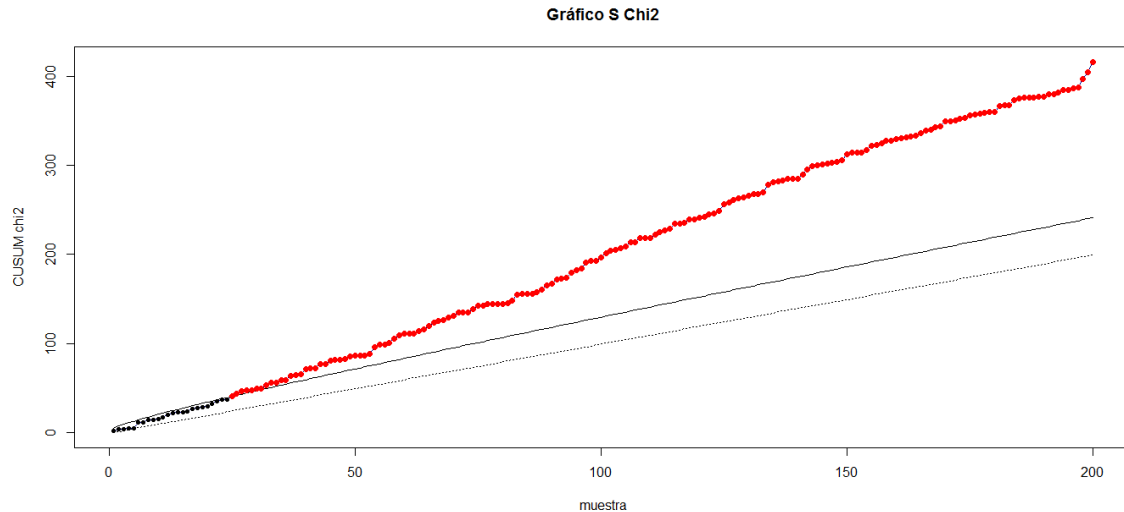
$$qchisq(1 - r_i, df = 1)$$

Luego, para la representación de los valores se van sumando cada uno de los rangos modificados que se tienen hasta el momento.

De la misma manera que en las otras variaciones se realiza el recálculo del límite de control y del valor central para cada nuevo dato que se va obteniendo, dicho cálculo se realiza mediante las siguientes expresiones:

$$LC = qchisq(1 - \alpha, df = i)$$

$$central = qchisq(0.5, df = i)$$



Comandos:
`graf.controlSChi2(new.rank,alpha=0.025)`
`rankd(new.depth,hist.depth)`
`depth.spatial(new.data, hist.data)`

Fuera de control muestras 25...

Figura 9: “Gráfico de control S Chi-cuadrado $N(1, 1)$ ”

4.2.4 DIFERENCIAS ENTRE GRÁFICOS

Al igual que para los gráficos Q, los puntos que se grafican pertenecen a una distribución normal multivariante con una unidad de diferencia respecto a los datos bajo control con los que se comparan. Al tratarse de los mismos datos que los gráficos Q, se observan 200 muestras.

En esta ocasión la diferencia entre los tres gráficos es mayor que en el caso anterior. En el gráfico que representa los promedios la rapidez de detección de la muestra como fuera de control se produce en la decimosexta muestra, en cambio para las otras dos gráficas hacen falta 9 muestras más para poder asegurar que ese proceso está fuera de control.

Cabe destacar que en el caso de promediar los rangos y en el de multiplicarlos antes de poder decir que el proceso se encuentra fuera de control y que hay cierta anomalía, se señalan previamente dos muestras como fuera de control. En cambio en el caso de la conversión de los rangos a una Chi-cuadrado la detección se produce para la primera muestra señalada como fuera de control. Esto indica que aunque la grafica de la Chi-cuadrado tarde más en dar un proceso como anómalo, el nivel de exactitud cuando lo hace es mucho mayor.

CAPÍTULO V

EJEMPLOS

5.1 DESCRIPCIÓN DE LAS SIMULACIONES

Las simulaciones que Liu [8] realizó se basan en la construcción de los gráficos de control tipo Q y S mediante el promedio de rangos usando el programa informático SUN workstation.

Ella generó la cantidad de 500 muestras históricas con la siguiente distribución $N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$ y posteriormente creó las nuevas muestras. Estas muestras son las que se han introducido en el programa y vienen detalladas en el anexo 4 de este documento. Las primeras 40 nuevas muestras siguen la misma distribución que las históricas y en las 40 siguientes se varía su distribución con el propósito de observar la rapidez de detección que presentan sus gráficos. La distribución que siguen los nuevos datos varía tanto la media como la desviación, por lo que va a haber una clara diferenciación con respecto a los históricos, los parámetros son los siguientes $N\left(\begin{pmatrix} 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix}\right)$.

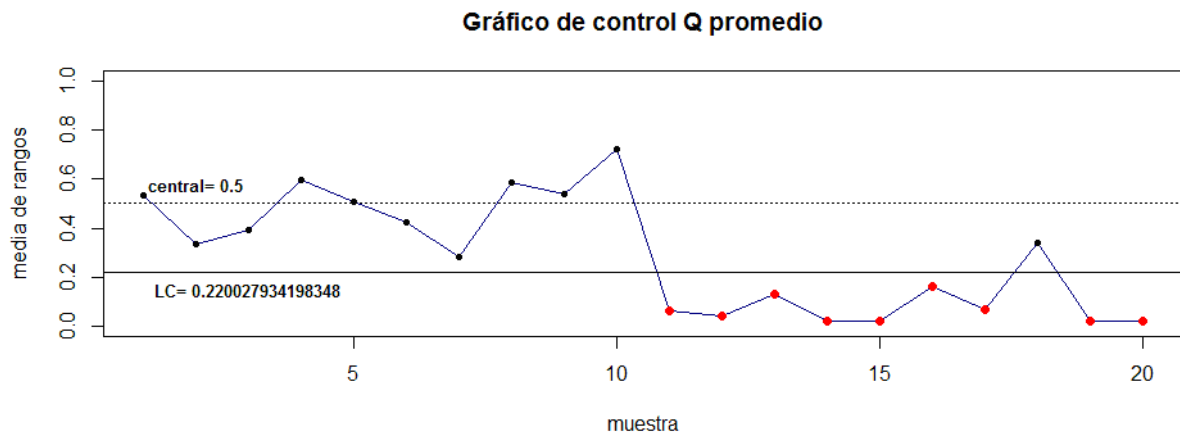
Como se ha explicado con anterioridad, la normalidad de las muestras no es un requisito necesario para el correcto funcionamiento de estos gráficos, pero en este caso se distribuyen de tal modo para facilitar los resultados.

Para la ordenación de los datos, Liu se basa en el concepto de profundidad que ya hemos detallado previamente, en este caso el tipo de profundidad que elige es la simplicial.

Este concepto de profundidad ya se ha explicado, pero además cabe decir que en comparación con la espacial, usada en este estudio, la profundidad simplicial es mucho más abrupta. Con esto se quiere decir que cuando puntos de la nube de datos están muy en la periferia se les asigna el valor 0 y se producen muchos empates, lo que supone la pérdida de información respecto a su ordenación. Por el contrario la profundidad espacial tiene una ordenación mucho más fina porque no se tiende a acotar de tal manera, a los puntos muy periféricos se les dan valores positivos pequeños.

A continuación se representan las simulaciones que implementa Liu con el programa diseñado. En este estudio los límites de control establecidos son los mismos que se han detallado antes, aunque no son exactamente los mismos que establece Liu, son una aproximación. De igual modo se ha escogido un tamaño muestral de 4.

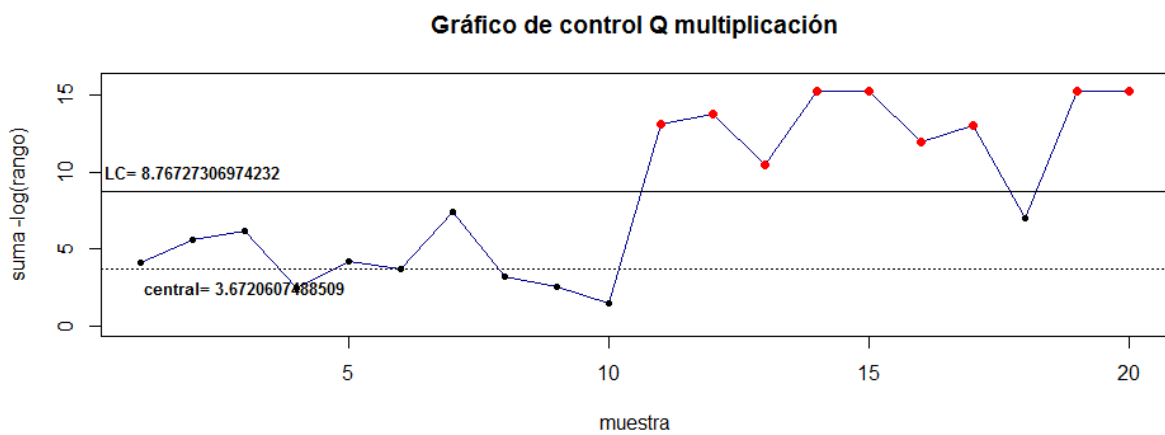
En los gráficos únicamente se representan las 40 muestras con la distribución de los datos históricos, las cuales son desde la 1 hasta la 40 y las 40 muestras con la anomalía que van desde la 41 hasta la 80.



Comandos:
`graf.controlQAv(Liu.ranks,alpha,n=4)`

Fuera de control muestras 11...

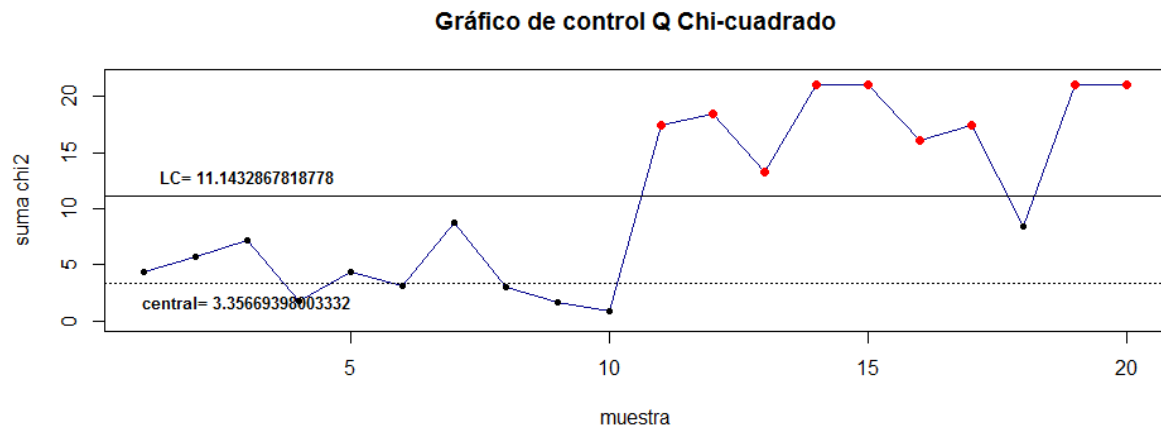
Figura 10: "Gráfico de control Q promedio rangos Liu"



Comandos:
`graf.controlQProd(Liu.ranks,alpha,n=4)`

Fuera de control muestras 11...

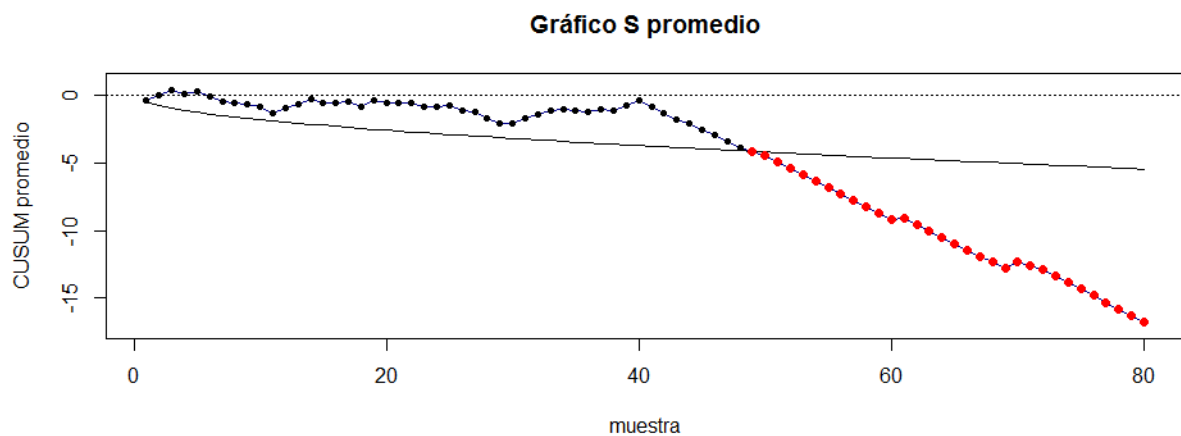
Figura 11: "Gráfico de control Q producto rangos Liu"



Comandos:
`graf.controlQChi2(Liu.ranks,alpha,n=4)`

Fuera de control muestras 11...

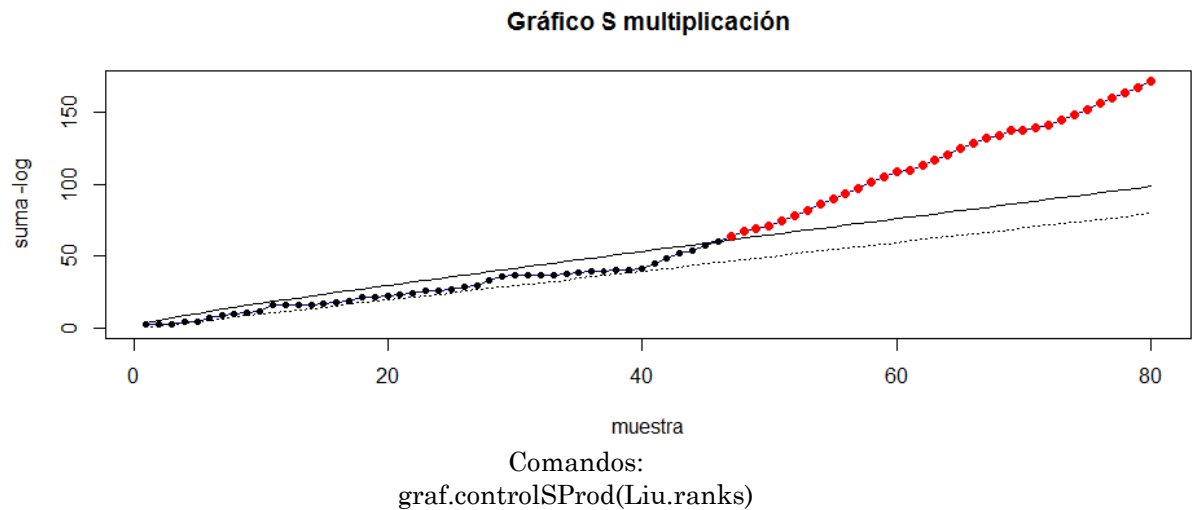
Figura 12: “Gráfico de control Q Chi-cuadrado rangos Liu”



Comandos:
`graf.controlSAv(Liu.ranks,k=500)`

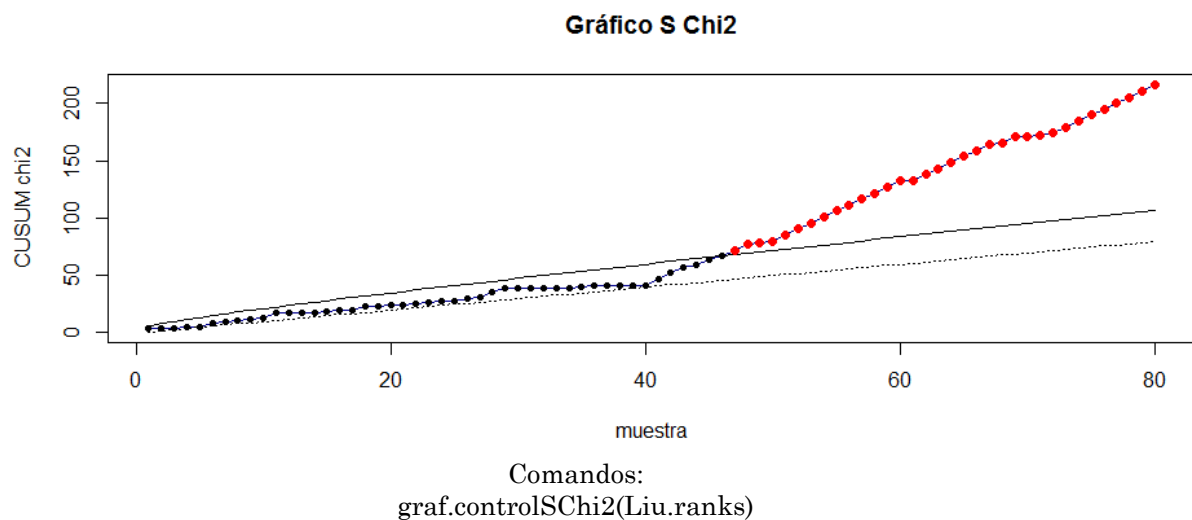
Fuera de control muestras 49...

Figura 13: “Gráfico de control S promedio rangos Liu”



Fuera de control muestras 47...

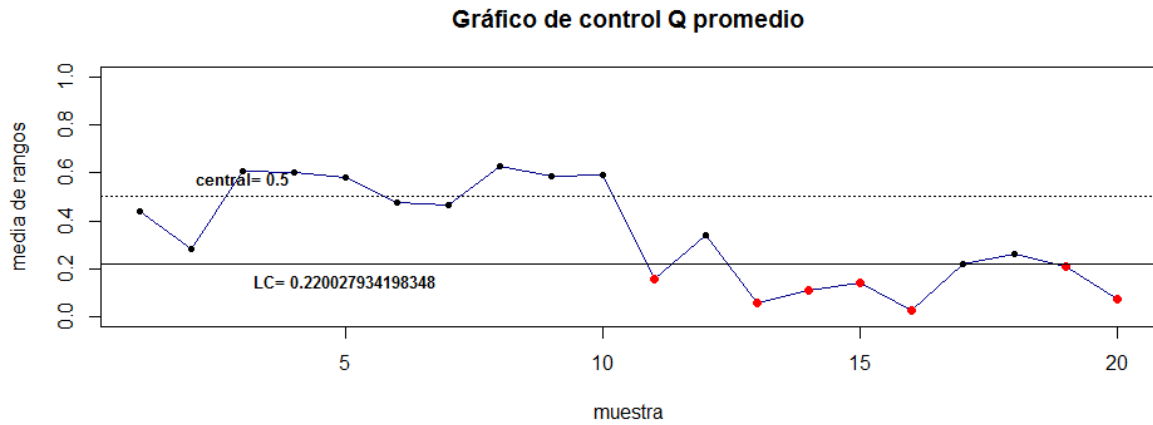
Figura 14: “Gráfico de control S producto rangos Liu”



Fuera de control muestras 47...

Figura 15: “Gráfico de control S Chi-cuadrado rangos Liu”

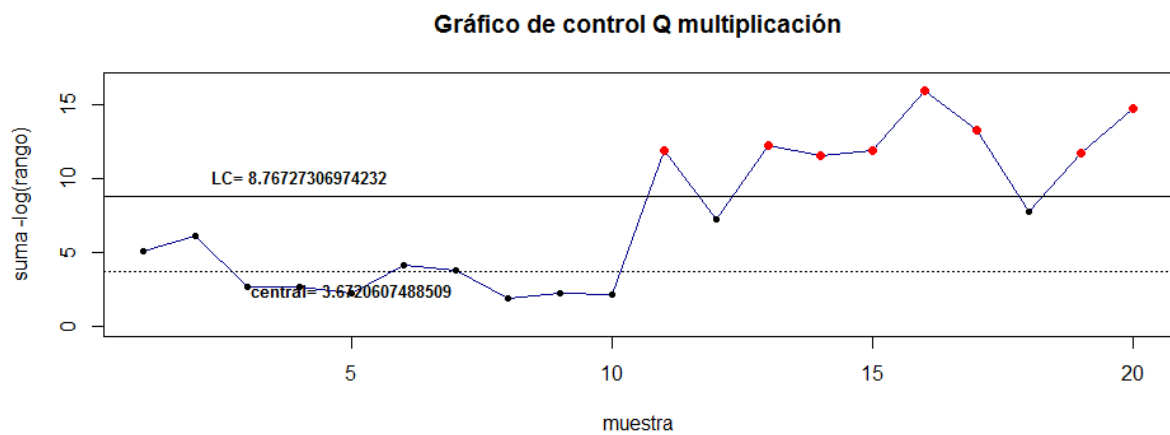
Ahora se representan una serie de datos distribuidos de la misma manera que los de Liu pero cuyos valores son totalmente aleatorios, generados por la profundidad espacial.



Comandos:
`graf.controlQAv(new.rank,n=4)`
`rankd(new.depth,hist.depth)`
`depth.spatial(new.data, hist.data)`

Fuera de control muestras 11, 13...

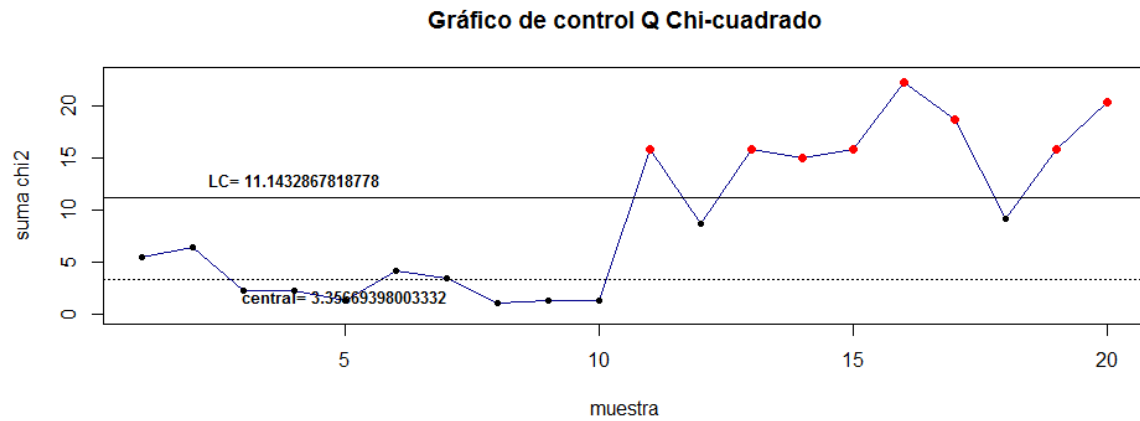
Figura 16: “Gráfico de control Q promedio $N(2, 4)$ ”



Comandos:
`graf.controlQProd(new.rank,n=4)`
`rankd(new.depth,hist.depth)`
`depth.spatial(new.data, hist.data)`

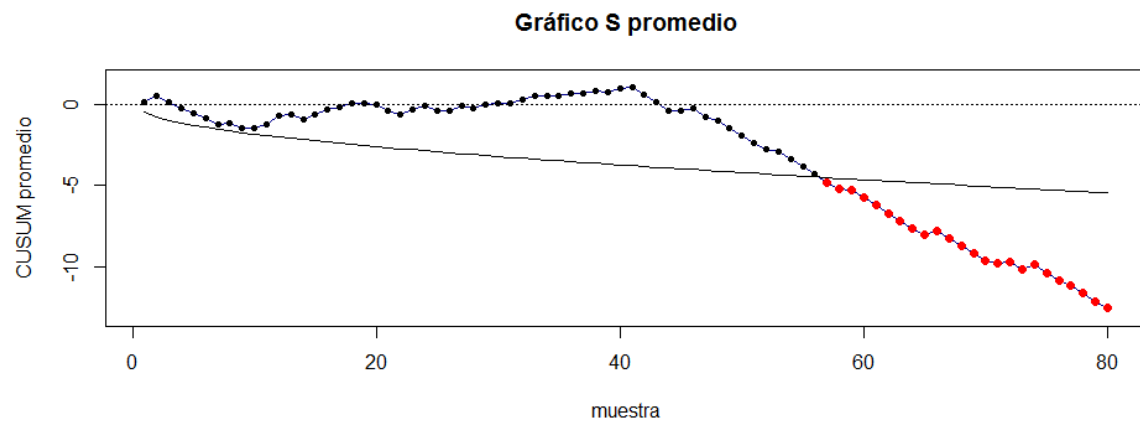
Fuera de control muestras 11, 13...

Figura 17: “Gráfico de control Q producto $N(2, 4)$ ”



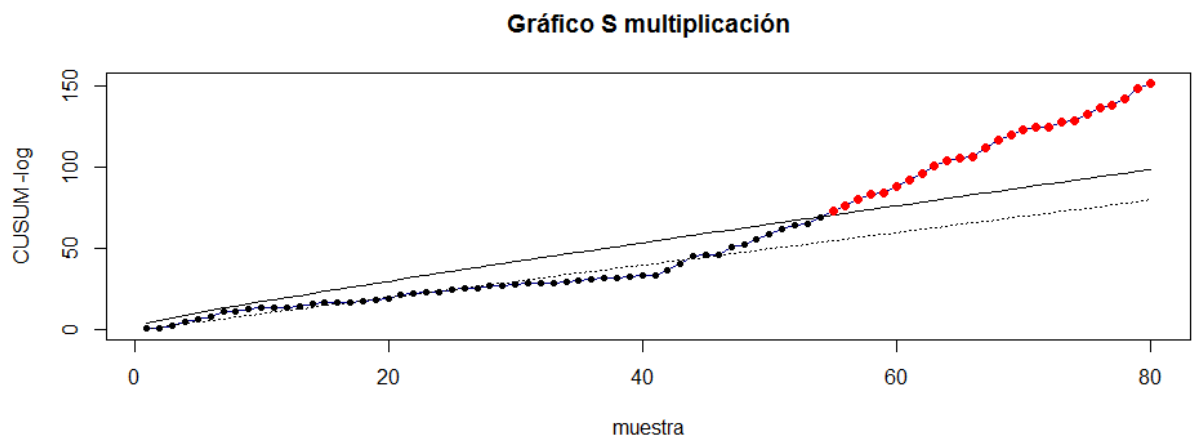
Fuera de control muestras 11, 13...

Figura 18: “Gráfico de control Q Chi-cuadrado $N(2, 4)$ ”



Fuera de control muestras 55...

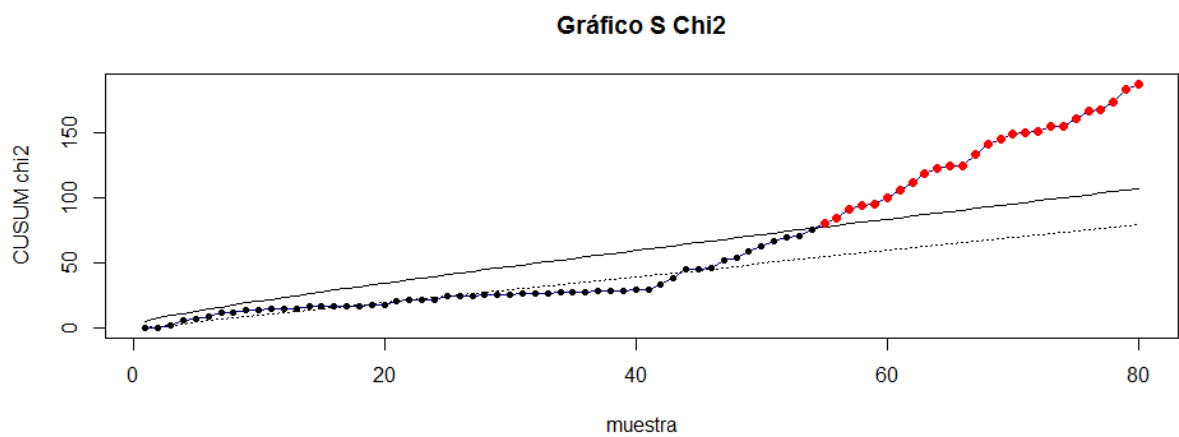
Figura 19: “Gráfico de control S promedio $N(2, 4)$ ”



Comandos:
`graf.controlSProd(new.rank,n=4)`
`rankd(new.depth,hist.depth)`
`depth.spatial(new.data, hist.data)`

Fuera de control muestras 57...

Figura 20: “Gráfico de control S producto $N(2, 4)$ ”



Comandos:
`graf.controlSChi2(new.rank,n=4)`
`rankd(new.depth,hist.depth)`
`depth.spatial(new.data, hist.data)`

Fuera de control muestras 57...

Figura 21: “Gráfico de control S Chi-cuadrado $N(2, 4)$ ”

5.2 COMENTARIOS DE LOS GRÁFICOS

Los resultados que ofrecen los gráficos Q son prácticamente idénticos, todos detectan la muestra como fuera de control en la primera muestra anómala.

Como caso especial en los gráficos de Liu está la muestra decimoctava la cual aparece como bajo control aun sabiendo que es anómala, este suceso también ocurre en la simulación propia de Liu, por lo que no es preocupante ni puede ser achacado al cambio en el tipo de profundidad entre ambas representaciones.

Aunque cabe señalar el hecho de que en los gráficos para los valores generados aleatoriamente, hay dos muestras marcadas como bajo control sin serlo e incluso en el caso del promedio existe una tercera.

Esto podría deberse a las diferencias entre profundidades de ambos gráficos.

En los gráficos S con los datos de Liu, si se puede ver cierta diferencia entre el que trabaja promediando los rangos y los otros dos. Este primero tarda 9 muestras en detectar el proceso como fuera de control, mientras que los que trabajan haciendo el producto y convirtiendo en Chi-cuadrado únicamente tardan 7 muestras.

Para el caso de los valores aleatorios, cabe decir que se comportan mucho peor que los anteriores por el hecho de que el proceso es detectado como anómalo en la quinceava muestra para el caso del producto y la transformación en Chi-cuadrado y en la decimoséptima para el caso del promedio.

Aunque el comportamiento para los datos aleatorios es peor, en general no hay diferencias en cuanto al uso de uno u otro tipo de transformación de los rangos.

Además en ninguno de estos casos se han detectado las muestras antes señaladas erróneamente como bajo control, por lo que se puede decir que los gráficos acumulativos aunque son menos rápidos a la hora de dar un proceso como fuera de control, lo hacen con una mayor eficiencia y en ellos no suelen aparecer muestras anómalas como controladas.

CAPÍTULO VI

COMPARATIVA

6.1 USO DE EXPERIMENTOS

La realización de experimentos es muy eficaz a la hora de reducir la variabilidad de los procesos.

Las simulaciones consisten básicamente en la variación de manera sistemática de los parámetros de la distribución de los datos que se están estudiando para ver la influencia que cada uno de ellos tiene en el resultado final del proceso.

En esta sección, se han realizado una serie de simulaciones para comparar la bondad de los tres tipos de gráficos Q, promedio, multiplicación y Chi-cuadrado, que se han detallado con anterioridad.

6.2 DESCRIPCIÓN DE LAS SIMULACIONES

Las simulaciones consisten en tomar una cantidad de datos históricos con parámetros conocidos $N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$. De la misma manera, se genera otra cantidad, 400 y 1000 en este caso, de nuevas observaciones en las que vamos a ir variando primeramente la media de cada una de ellas y posteriormente lo haremos con la desviación estándar.

El experimento consiste en comparar los rangos de estas nuevas observaciones con el límite de control establecido para cada tipo de gráfico, de igual modo que en los gráficos que se han explicado previamente, se van a coger muestras de tamaño 4.

El programa creado se encarga de guardar el número de muestras que han sido simuladas hasta detectar una anomalía y posteriormente se promedian hasta la detección para cada conjunto de datos históricos.

Se realizan variaciones escalonadas y pequeñas que permiten ver como varía la cantidad de muestras simuladas hasta la anomalía de manera clara.

Al igual que se ha hecho a la hora de implementar los gráficos anteriores y siguiendo los experimentos de Liu, se establece un valor de $\alpha=0.025$

Este valor de α está bastante alejado del que comúnmente se suele utilizar en estadística (0.0027). Esto supone que se va a obtener mayor cantidad de anomalías y por tanto, muchas de ellas pueden ser falsas alarmas, lo que va a indicar es que se trabaja con situaciones mucho más críticas.

A la hora de comparar se buscan resultados que se puedan interpretar fácilmente. Como se ha comentado con anterioridad, para ver la rapidez con la que se encuentra una anomalía no se va a representar mediante el valor de alpha el cual indica la probabilidad de que cualquiera de los datos que han sido simulados sobrepase los límites de control, sino que se hace con el parámetro ARL, del inglés *Average Run Length*, que indica para este estudio la cantidad de muestras que han sido simuladas hasta detectar una anomalía y se puede hallar mediante la expresión $1/\alpha$.

La cantidad de datos históricos que se tienen influye en la bondad de cada gráfico es decir, que pueda dar buenas aproximaciones. Liu indica en sus experimentos [8] que para 40 observaciones, los resultados apenas son visibles, por lo que para este proyecto se ha decidido hacerlos para 400 y 1000 observaciones.

6.2.1 VARIACIÓN DE LA MEDIA

Se ha decidido únicamente variar una coordenada de la media para que sea más fácil de visualizar.

El proceso consiste en la generación de nuevas muestras hasta que el programa detecte por primera vez una muestra fuera de control, cuando esto ocurre se pasa de manera automática al siguiente valor que puede tomar la media.

Los valores que va a ir tomando la coordenada x de la media son los siguientes: 0, 0.25, 0.5, 0.75, 1, 1.5 y 2.

Los resultados obtenidos tras las simulaciones para una cantidad de datos históricos de 1000 y siguiendo cada una de las modificaciones del parámetro λ (lambda), el cual representa la coordenada x de la media para la generación de las nuevas observaciones, es decir que siguen esta distribución $N\left(\begin{pmatrix} \lambda \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$, son los siguientes:

λ	Promedio		Producto		Chi-cuadrado	
	media	desviación	media	desviación	media	desviación
0	41.232	42.732	42.102	46.533	42.236	42.983
0.25	34.176	34.003	35.032	37.124	35.683	36.597
0.5	23.591	23.862	22.483	23.445	23.395	23.985
0.75	13.845	14.657	12.719	12.516	12.614	12.627
1	8.243	8.21	6.936	6.77	7	6.766
1.5	2.856	2.269	2.566	2.082	2.455	1.915
2	1.549	0.916	1.358	0.707	1.412	0.783

Figura 22: “Tabla comparativa variación en la media (1000 datos históricos)”

Se ve que cuando no existe diferenciación alguna entre los parámetros de los datos históricos y los nuevos, el gráfico Q para el promedio es aproximadamente 40, por lo que se asemeja bastante bien a los resultados ideales que se suponía que se iban a dar. Esto quiere decir que aunque se esté trabajando con muestras cuyos parámetros son los mismos y el gráfico esté bajo control, se generan ciertas muestras anómalas, cuya frecuencia viene predefinida por el valor de alpha que tomemos, como en este estudio es de 0.025, se obtiene una muestra anómala aproximadamente cada 40 observaciones ($1/0.025=40$).

En cambio, para los gráficos del producto y la Chi-cuadrado los resultados varían un poco más, por lo que se puede decir que el ajuste que hacen respecto a la distribución original no es tan bueno.

Para valores de λ grandes se puede observar que los gráficos basados en el producto y la conversión en Chi-cuadrado de los rangos funcionan mucho mejor que el del promedio. Aunque también es verdad que esa diferenciación en cuanto a resultados no es tan

marcada como en el caso anterior, ya que ahora se está diferenciando entre tener valores del ARL entorno a 1.4 para el promedio al 1.6 del producto.

A continuación se muestran los distintos resultados que se obtienen al variar el parámetro lambda cuando la cantidad de observaciones históricas es de 400:

λ	Promedio		Producto		Chi-cuadrado	
	media	desviación	media	desviación	media	desviación
0	40.69	40.851	45.396	48.7	51.663	61.786
0.25	35.725	34.674	40.413	45.709	41.99	49.502
0.5	24.154	27.884	26.056	29.374	28.029	32.382
0.75	15.028	14.632	13.661	14.695	13.69	14.942
1	8.265	8.191	7.436	7.308	7.461	7.323
1.5	2.773	2.343	2.475	1.913	2.728	2.421
2	1.59	0.956	1.4	0.757	1.395	0.772

Figura 23: “Tabla comparativa variación en la media (400 datos históricos)”

Aquí se puede observar con claridad, que cuando no se dispone de gran cantidad de observaciones es mejor no utilizar los gráficos en los que los rangos se multiplican y se convierten en Chi-cuadrado, ya que su variación en comparación con el ideal es muy dispar, por el contrario se ve un ajuste muy bueno en el gráfico que promedia los rangos cuando no se ha introducido ninguna variación en la media de las nuevas observaciones que se generan.

Aunque si la variación que se produce en la media es grande, los gráficos del producto y de la Chi-cuadrado detectan un poco mejor que el del promedio, entorno a los valores que se han obtenido en el experimento anterior.

Se puede concluir que la mejor o peor elección de cada una de las gráficas depende claramente del uso que le vayamos a dar y la cantidad de observaciones de las que se disponga.

Hay que señalar que mayoritariamente se busca que los gráficos de control nos proporcionen con seguridad cuando una muestra con un claro desajuste está el proceso. Por lo que para estas ocasiones es mejor utilizar los gráficos del producto y conversión en Chi-cuadrado de los rangos, aunque la diferencia con los resultados del gráfico que promedia es mínima.

Pero bien es cierto que hay en determinadas ocasiones que lo que se busca es la precisión, por lo que en estos casos será recomendable utilizar el gráfico del promedio, ya que para variaciones pequeñas es el que mejor resultado ha dado. Aunque la mejor solución para estos casos sería el uso de los gráficos con memoria, los cuales ya se han explicado anteriormente.

6.2.2 VARIACIÓN DE LA DESVIACIÓN ESTÁNDAR

Para esta serie de experimentos se va a ir variando la desviación de los nuevos datos que se van generando, es decir la dispersión respecto al punto central de generación va a ir en aumento.

Ocurre todo de la misma manera que para el caso de las variaciones en la media, se van generando nuevos datos hasta que se detecta una anomalía y se van modificando el parámetro de la desviación.

En este caso los valores que va a tomar la desviación típica son: 1, 1.25, 1.5, 1.75, 2 y 2.5

La siguiente tabla muestra los resultados para cada uno de los valores de la desviación que se han simulado, cuando la cantidad de observaciones históricas es de 1000. En esta ocasión esta es la distribución que siguen $N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \lambda^2 & 0 \\ 0 & \lambda^2 \end{pmatrix}\right)$:

λ	Promedio		Producto		Chi-cuadrado	
	media	desviación	media	desviación	media	desviación
1	41.227	42.735	42.477	43.301	42.523	46.568
1.25	7.471	6.972	5.8	5.467	5.633	5.444
1.5	3.364	2.813	2.265	1.705	2.194	1.718
1.75	2.165	1.653	1.509	0.905	1.463	0.783
2	1.693	1.119	1.248	0.558	1.226	0.507
2.5	1.323	0.658	1.049	0.216	1.058	0.246

Figura 24: “Tabla comparativa variación en la desviación (1000 datos históricos)”

A la vista de estos resultados, se puede observar que los gráficos en los que se varía el rango mediante su multiplicación y en los que se transforman en Chi-cuadrado proporcionan un ajuste similar.

Aunque sigue siendo mejor el resultado del ARL que ofrece el gráfico del promedio, los otros dos son bastante similares.

También se puede observar que en este caso la desviación de los datos influye en gran medida en los resultados, ya que las variaciones que se realizan en la desviación son pequeñas y la disminución del ARL es considerable, sobre todo entre los valores de 1 y 1.25, por lo que los resultados que ofrecen los gráficos de control cuando suceden estas variaciones no son para nada fiables.

A continuación se expone para una cantidad de observaciones históricas de 400, los diferentes resultados obtenidos con cada uno de los valores de la desviación típica indicados:

λ	Promedio		Producto		Chi-cuadrado	
	media	desviación	media	desviación	media	desviación
1	40.443	42.337	46.664	52.006	51.083	54.625
1.25	7.603	7.308	5.989	5.445	6.066	6.153
1.5	3.586	2.912	2.404	1.744	2.35	1.805
1.75	2.122	1.573	1.549	0.915	1.544	0.966
2	1.719	1.124	1.285	0.615	1.261	0.577
2.5	1.316	0.653	1.084	0.302	1.063	0.247

Figura 25: “Tabla comparativa variación en la desviación (400 datos históricos)”

Para este caso en el que la cantidad de muestras históricas es menor, se puede ver que las opciones del producto y de la Chi-cuadrado son poco recomendables debido a su poca

cercanía con el valor ideal de 40 para el ARL, en cambio el gráfico que promedia tiene un ajuste muy bueno.

En cuanto a la diferencia del ARL cuando se producen las variaciones en las desviaciones típicas, aproximadamente ocurre lo mismo que en el caso anterior. Un gran salto cuando se realiza la primera variación, pasando de 1 a 1.25 y luego se ve la disminución del ARL de manera escalonada.

Para finalizar se puede concluir que la cantidad de datos históricos que se tengan en un experimento de este tipo apenas influye en el ARL si se va a usar el gráfico del promedio, que es el más recomendable sobre todo si no se produce variación.

En cuanto cambia la desviación típica de los datos los resultados de los gráficos son muy similares y el tener mayor o menor cantidad de históricos es irrelevante.

CAPÍTULO VII

CONCLUSIONES

7.1 CONCLUSIONES

En el desarrollo del proyecto se ha podido ver los diferentes tipos de gráficos más destacados que hay para el estudio del control estadístico de procesos. Centrándose en los procesos multivariantes no paramétricos y por tanto en los gráficos Q y S, de los cuales se han ido presentado en todo momento comparaciones entre ambos así como entre las diferentes maneras de tratar los rangos de las observaciones.

Se trataba de conocer la bondad de los diferentes gráficos que detalló Liu para observaciones multivariantes no paramétricas. Aunque solo se han realizado experimentos con distribuciones normales, se sabe que también funcionan cuando la normalidad no está presente.

También se ha ahondado en el concepto de profundidad y se ha señalado la importancia de escoger uno u otro según la función que se le vaya a dar. Con la implementación de los gráficos con los valores que Liu tomó se ha podido ver las diferencias existentes entre el uso de las distintas profundidades más claramente.

Con este estudio se buscaba conocer qué gráfico es el que mejor detecta anomalías para las observaciones multivariantes. Finalmente se ha llegado a la conclusión de que no hay un único gráfico que se pueda determinar como el mejor. Todo depende del uso que se vaya a hacer y de las condiciones en las que se desarrolle.

Así se ha podido ver que las transformaciones de los rangos mediante su producto o su conversión a Chi-cuadrado dan más importancia a las observaciones extremas, es decir aquellas realmente alejadas, y por lo tanto su funcionamiento es mejor que el del promedio. Esto ocurre tanto cuando la cantidad de datos históricos es 1000 como cuando es 400. Aunque hay que señalar que cuando la cantidad es de 400 el ARL_0 se controla muy mal para los casos de la Chi-cuadrado y del producto, por lo que habría que realizar algún refinamiento.

Por lo tanto para las pequeñas desviaciones es mejor el uso del promedio.

Para los casos en los que varía la desviación estándar el cambio que se observa es muy brusco, debido a que lo que se está haciendo es aumentar la dispersión y por lo tanto las observaciones están en la zona periférica con lo que la detección de anomalías es más propensa.

Por último se ha podido observar con claridad la diferencia entre los gráficos Q y S, el gráfico Q alerta mucho antes de que el proceso se ha descontrolado, pero hay que prestar mucha atención ya que en ocasiones algunas observaciones aparecen como controladas y en realidad no lo están.

Sin embargo la eficacia del gráfico S es mucho mayor, cuando alerta de anomalías no suele fallar y no da como buena alguna que no lo sea, pero la rapidez en comparación con el Q a la hora de detectarlas es mucho menor.

Se podría decir que este último es un gráfico más conservador y más preciso, debido también a que tiene mucho más en cuenta las pequeñas distorsiones que se producen en cada una de las observaciones, ya que de manera contraria al Q, toma las observaciones de una a una y no hace agrupaciones.

Bibliografía

- [1] César Pérez (1998). *Control estadístico de la calidad*. Rama Editorial, Madrid
- [2] Douglas C. Montgomery (2004). *Control estadístico de la calidad*, 3ª ed., Wiley, Nueva York.
- [3] Emmanuel Paradis. *R para Principiantes* (2003). Disponible en Web: http://cran.r-project.org/doc/contrib/rdebuts_es.pdf
- [4] Ignacio Cascos, Ángel López y Juan Romo (2011). Data depth in Multivariate Statistics. *Boletín de Estadística e Investigación Operativa*. Vol. 27, No. 3, pp. 151-174. Disponible en Web: http://www.seio.es/BEIO/files/BEIOVol27Num3_Estadistica-Web.pdf
- [5] John W. Tukey (1977). *Exploratory Data Analysis*. Addison-Wesley, Reading, MA.
- [6] Julio S. Santana y Efraín Mateos. *El arte de programar en R: un lenguaje para la estadística* (2014). Disponible en Web: https://cran.r-project.org/doc/contrib/Santana_El_arte_de_programar_en_R.pdf
- [7] Prasanta C. Mahalanobis (1936). *On the generalised distance in statistics*. Volume 2, pages 49-55.
- [8] Regina Y. Liu (1995). Control charts for multivariate processes. *Journal of the American Statistical Association*. Vol. 90, No 432, Theory and Methods.
- [9] Regina Y. Liu (1990). On a notion of data depth based on random simplices. *Ann. Statist.*, 18:405-414.
- [10] Yehuda Vardi y Cun-Hui Zhang (2000). The multivariate L_1 -median and associated data depth. *Proceedings of the National Academy of Sciences of the United States of America*. Vol. 97, No. 4, pp. 1423-1426.

Anexos

1. DETALLE DE LAS FUNCIONES

Rankd(new,hist)

Input: new- vector de profundidades de los nuevos datos.
hist- vector de profundidades de los datos históricos.

Output: vector de rangos de los nuevos datos calculado mediante las profundidades.

graf.controlQAv(rank,n,alpha)

Input: rank- vector de rangos.
n- tamaño muestral.
alpha- nivel de significación.

Output: representa el gráfico de control Q en el cual los rangos se promedian.

graf.controlQProd(rank,n,alpha)

Input: rank- vector de rangos.
n- tamaño muestral.
alpha- nivel de significación.

Output: representa el gráfico de control Q en el cual los rangos se multiplican.

graf.controlQChi2 (rank,n,alpha)

Input: rank- vector de rangos.
n- tamaño muestral.
alpha- nivel de significación.

Output: representa el gráfico de control Q en el cual los rangos se convierten a una Chi-cuadrado.

graf.controlSAv(rank,n,alpha)

Input: rank- vector de rangos.
n- tamaño muestral.
alpha- nivel de significación.

Output: representa el gráfico de control S en el cual los rangos se promedian.

graf.controlSProd(rank,n,alpha)

Input: rank- vector de rangos.
n- tamaño muestral.
alpha- nivel de significación.

Output: representa el gráfico de control S en el cual los rangos se multiplican.

graf.controlSChi2(rank,n,alpha)

Input: rank- vector de rangos.
n- tamaño muestral.
alpha- nivel de significación.

Output: representa el gráfico de control S en el cual los rangos se convierten a Chi-cuadrado.

2. CÓDIGOS EN R UTILIZADOS

- Función para la obtención de los rangos de los nuevos datos:

Rango.R

```
rankd<-function(new,hist){  
  m<-length(hist)  
  v<-length(new)  
  rank<-vector(length=v)  
  for(i in 1:v){  
    rank[i]<-sum(hist<=new[i])+sum(new<=new[i])  
  }  
  return(rank/(m+v))  
}
```

- Funciones para la representación de cada uno de los tipos de gráficos Q detallados:

Graf.ControlQAv.R

```
graf.controlQAv<-function(rank,n,alpha=0.025){  
  
  m<-matrix(rank,ncol=n,byrow=T)  
  central<-0.5  
  divX<-1:nrow(m)  
  op<-apply(m, 1, sum)  
  medias<-op/n  
  CL<-((factorial(n)*alpha)^(1/n))/n  
  
  win.graph(width=10,height=4)  
  plot(medias,ylim=c(0,1),col="darkblue",type="l",main="Gráfico de control Q  
promedio",ylab="media de rangos",xlab="muestra")  
  
  d<-length(medias)  
  
  for(i in 1:d){  
    if(medias[i]<CL){  
      points(divX[i],medias[i],col='red',pch=16)  
      print('muestra fuera de control')  
      print(divX[i])  
    }  
    else{  
      points(divX[i],medias[i],col='black',pch=20)  
    }  
  }  
  abline(h=central,lty=3)  
  text((min(divX)+2),central,paste('central=',central),pos=3,font=2,cex=0.8)  
  abline(h=CL)  
  text((min(divX)+4),CL,paste('LC=',CL),pos=1,font=2,cex=0.8)  
}
```


Graf.ControlQProd.R

```
graf.controlQProd<-function(rank,n,alpha=0.025){

  m<-matrix(-log(rank),ncol=n,byrow=T)
  central<-qgamma(0.5, shape=n, rate=1)
  divX<-1:nrow(m)
  op<-apply(m, 1, sum)
  d<-length(op)
  CL<-qgamma(1-alpha, shape=n, rate=1)

  win.graph(width=10,height=4)
  plot(op,type="l",col="darkblue",ylim=c(0,max(op,CL)+0.5),main="Gráfico de
  control Q multiplicación",ylab="suma -log(rango)",xlab="muestra")

  for(i in 1:d){
    if(op[i]>CL){
      points(divX[i],op[i],col='red',pch=16)
      print('muestra fuera de control')
      print(divX[i])
    }
    else{
      points(divX[i],op[i],col='black',pch=20)
    }
  }
  abline(h=central,lty=3)
  text((min(divX)+4),central,paste('central=',central),pos=1,font=2,cex=0.8)
  abline(h=CL)
  text((min(divX)+3),CL,paste('LC=',CL),pos=3,font=2,cex=0.8)

}
```

Graf.ControlQChi2.R

```
graf.controlQchi2<-function(rank,n,alpha=0.025){

  m<-matrix(qchisq(1-rank,df=1),ncol=n,byrow=T)
  central<-qchisq(0.5,df=n)
  divX<-1:nrow(m)
  op<-apply(m, 1, sum)
  CL<-qchisq(1-alpha,df=n)

  win.graph(width=10,height=4)
  plot(op,ylim=c(0,max(op,CL)+0.5),col="darkblue",type="l",main="Gráfico de
  control Q Chi-cuadrado",ylab="suma chi2",xlab="muestra")
  points(CL,type="l",col='blue')

  d<-length(op)

  for(i in 1:d){
    if(op[i]>CL){
      points(divX[i],op[i],col='red',pch=16)
      print('muestra fuera de control')
      print(divX[i])
    }
    else{
      points(divX[i],op[i],col='black',pch=20)
    }
  }
  abline(h=central,lty=3)
  text((min(divX)+4),central,paste('central=',central),pos=1,font=2,cex=0.8)
  abline(h=CL)
  text((min(divX)+3),CL,paste('LC=',CL),pos=3,font=2,cex=0.8)

}
```

- Funciones para la representación de cada uno de los tipos de gráficos S detallados:

Graf ControlSAv.R

```
graf.controlSAv<-function(rank,k,alpha=0.025){
  d<-length(rank)
  v=vector(length=d)
  v<-(rank-0.5)
  for(i in 2:d){
    v[i]=v[i]+v[i-1]
  }

  central<-0
  CL=vector(length=d)

  for(i in 1:d){
    if(alpha<=1/factorial(i)){
      CL[i]=(((factorial(i)*alpha)^(1/i))-i/2)
    }
    else{
      CL[i]=qnorm(alpha)*i*sqrt((1/k+1/i)/12)
    }
  }

  win.graph(width=10,height=4)
  plot(v,ylim=c((min(v,CL).5),(max(v,CL)+.5)),col="darkblue",type="l",main=
"Gráfico S promedio",ylab="CUSUM promedio",xlab="muestra")
  points(CL,type="l")

  for(i in 1:d){
    if(v[i]<CL[i]){
      points(i,v[i],col='red',pch=16)
      print('muestra fuera de control')
      print(i)
    }
    else{
      points(i,v[i],col='black',pch=20)
    }
  }
  abline(h=central,lty=3)
}
```

Graf ControlSProd.R

```
graf.controlSProd<-function(rank,alpha=0.025){
  v<-(-log(rank))
  d<-length(rank)
  central=vector(length=d)
  CL=vector(length=d)
  for(i in 1:d){
    CL[i]<-qgamma(1-alpha, shape=i, rate=1)
    central[i]<-qgamma(0.5, shape=i, rate=1)
  }

  for(i in 1:d){
    if(i>1){
      v[i]=v[i-1]+v[i]
    }
    else{
      v[i]=v[i]
    }
  }
}
```

```

    }
  }

  win.graph(width=10,height=4)
  plot(v,ylim=c(0,max(v,CL)+.5),col="darkblue",type="l",main="Gráfico S
multiplicación",ylab="CUSUM -log",xlab="muestra")
  points(CL,type="l")
  lines(central,lty=3)

  for(i in 1:d){
    if(v[i]>CL[i]){
      points(i,v[i],col='red',pch=16)
      print('muestra fuera de control')
      print(i)
    }
    else{
      points(i,v[i],col='black',pch=20)
    }
  }
}

```

Gráf ControlSchi2.R

```

graf.controlSchi2<-function(rank,alpha=0.025){
  v<-qchisq(1-rank,df=1)
  d<-length(rank)

  for(i in 2:d){
    if(i>1){
      v[i]=v[i-1]+v[i]
    }
  }

  CL=vector(length=d)
  central=vector(length=d)

  for(i in 1:d){
    CL[i]=qchisq(1-alpha,df=i)
    central[i]=qchisq(0.5,df=i)
  }

  win.graph(width=10,height=4)
  plot(v,ylim=c(0,max(v,CL)+0.5),col="darkblue",type="l",main="Gráfico S
Chi2",ylab="CUSUM chi2",xlab="muestra")
  points(CL,type="l")
  lines(central,lty=3)

  for(i in 1:d){
    if(v[i]>CL[i]){
      points(i,v[i],col='red',pch=16)
      print('muestra fuera de control')
      print(i)
    }
    else{
      points(i,v[i],col='black',pch=20)
    }
  }
}

```

- Códigos utilizados para la realización de las comparativas del capítulo VI:

Simulacion Av.R

```

m=1000
h=1000 #Toma el valor 400 para la otra simulación
alpha=0.025
n=4
nmuestras=vector(length=m)
tot.sd=vector(length=7)
tot.av=vector(length=7)

lambda=c(0,.25,.5,.75,1,1.5,2)

for(j in 1:7) {
  for(i in 1:m) {
    hist.data<-mvrnorm(h, rep(0,2),matrix(c(1,0,0,1),nrow = 2))
    hist.depth<-depth.spatial(hist.data,hist.data)
    nmuestras[i]=0
    anom=F
    while(anom==F){
      new.data<-mvrnorm(n,c(lambda[j],0),matrix(c(1,0,0,1),nrow = 2))
      nmuestras[i]=nmuestras[i]+1
      new.depth<-depth.spatial(new.data,hist.data)
      new.rank<-rankd(new.depth,hist.depth)

      CL=((factorial(n)*alpha)^(1/n))/n

      if(mean(new.rank)<CL){anom=T}
    }
  }

  tot.av[j]<-mean(nmuestras)
  tot.sd[j]<-sd(nmuestras)
}
tot.av
tot.sd

```

Simulacion Prod.R

```

m=1000
h=1000 #Toma el valor 400 para la otra simulación
alpha=0.025
n=4
nmuestras=vector(length=m)
tot.sd=vector(length=7)
tot.prod=vector(length=7)
lambda=c(0,.25,.5,.75,1,1.5,2)

for(j in 1:7) {
  for(i in 1:m) {
    hist.data<-mvrnorm(h, rep(0,2),matrix(c(1,0,0,1),nrow = 2))
    hist.depth<-depth.spatial(hist.data,hist.data)
    nmuestras[i]=0
    anom=F
    while(anom==F){
      new.data<-mvrnorm(n,c(lambda[j],0),matrix(c(1,0,0,1),nrow = 2))
      nmuestras[i]=nmuestras[i]+1
      new.depth<-depth.spatial(new.data,hist.data)
      new.rank<-rankd(new.depth,hist.depth)

      CL=qgamma(1-alpha, shape=n, rate=1)

      if(sum(-log(new.rank))>CL){anom=T}
    }
  }
}

```

```

    }
  }
  tot.prod[j]<-mean(nmuestras)
  tot.sd[j]<-sd(nmuestras)
}
tot.prod
tot.sd

```

Simulacion Chi2.R

```

m=1000
h=1000 #Toma el valor 400 para la otra simulación
alpha=0.025
n=4
tot.sd=vector(length=7)
tot.chi=vector(length=7)
nmuestras=vector(length=m)
lambda=c(0,.25,.5,.75,1,1.5,2)

for(j in 1:7) {
  for(i in 1:m) {
    hist.data<-mvrnorm(h, rep(0,2),matrix(c(1,0,0,1),nrow = 2))
    hist.depth<-depth.spatial(hist.data,hist.data)
    nmuestras[i]=0
    anom=F
    while(anom==F){
      new.data<-mvrnorm(n,c(lambda[j],0),matrix(c(1,0,0,1),nrow = 2))
      nmuestras[i]=nmuestras[i]+1
      new.depth<-depth.spatial(new.data,hist.data)
      new.rank<-rankd(new.depth,hist.depth)

      CL<-qchisq(1-alpha,df=n)

      if(sum(qchisq(1-new.rank,df=1))>CL){anom=T}
    }
  }
  tot.chi[j]<-mean(nmuestras)
  tot.sd[j]<-sd(nmuestras)
}
tot.chi
tot.sd

```

3. TABLA DE CONSTANTES PARA GRÁFICOS DE CONTROL

Constantes para Gráficos de Control																
n	A	A2	A3	c4	1/c4	B3	B4	B5	B6	d2	d3	1/d2	D1	D2	D3	D4
2	2.121	1.880	2.639	0.798	1.253	0.000	3.267	0.000	2.606	1.128	0.853	0.886	0.000	3.686	0.000	3.267
3	1.732	1.023	1.954	0.886	1.128	0.000	2.568	0.000	2.276	1.693	0.888	0.591	0.000	4.338	0.000	2.575
4	1.500	0.729	1.628	0.921	1.085	0.000	2.266	0.000	2.088	2.059	0.880	0.486	0.000	4.698	0.000	2.282
5	1.342	0.577	1.427	0.940	1.064	0.000	2.089	0.000	1.964	2.326	0.864	0.430	0.000	4.918	0.000	2.114
6	1.225	0.483	1.287	0.952	1.051	0.030	1.970	0.029	1.874	2.534	0.848	0.395	0.000	5.079	0.000	2.004
7	1.134	0.419	1.182	0.959	1.042	0.118	1.882	0.113	1.806	2.704	0.833	0.370	0.205	5.204	0.076	1.924
8	1.061	0.373	1.099	0.965	1.036	0.185	1.815	0.179	1.751	2.847	0.820	0.351	0.388	5.307	0.136	1.864
9	1.000	0.337	1.032	0.969	1.032	0.239	1.761	0.232	1.707	2.970	0.808	0.337	0.547	5.394	0.184	1.816
10	0.949	0.308	0.975	0.973	1.028	0.284	1.716	0.276	1.669	3.078	0.797	0.325	0.686	5.469	0.223	1.777
11	0.905	0.285	0.927	0.975	1.025	0.321	1.679	0.313	1.637	3.173	0.787	0.315	0.811	5.535	0.256	1.744
12	0.866	0.266	0.886	0.978	1.023	0.354	1.646	0.346	1.610	3.258	0.778	0.307	0.923	5.594	0.283	1.717
13	0.832	0.249	0.850	0.979	1.021	0.382	1.618	0.374	1.585	3.336	0.770	0.300	1.025	5.647	0.307	1.693
14	0.802	0.235	0.817	0.981	1.019	0.406	1.594	0.398	1.563	3.407	0.763	0.294	1.118	5.696	0.328	1.672
15	0.775	0.223	0.789	0.982	1.018	0.428	1.572	0.421	1.544	3.472	0.756	0.288	1.203	5.740	0.347	1.653
16	0.750	0.212	0.763	0.983	1.017	0.448	1.552	0.440	1.527	3.532	0.750	0.283	1.282	5.782	0.363	1.637
17	0.728	0.203	0.739	0.985	1.016	0.466	1.534	0.459	1.510	3.588	0.744	0.279	1.356	5.820	0.378	1.622
18	0.707	0.194	0.718	0.985	1.015	0.482	1.518	0.475	1.496	3.640	0.739	0.275	1.424	5.856	0.391	1.609
19	0.688	0.187	0.698	0.986	1.014	0.497	1.503	0.490	1.483	3.689	0.733	0.271	1.489	5.889	0.404	1.596
20	0.671	0.180	0.680	0.987	1.013	0.510	1.490	0.503	1.470	3.735	0.729	0.268	1.549	5.921	0.415	1.585
21	0.655	0.173	0.663	0.988	1.013	0.523	1.477	0.516	1.459	3.778	0.724	0.265	1.606	5.951	0.425	1.575
22	0.640	0.167	0.647	0.988	1.012	0.534	1.466	0.528	1.448	3.819	0.720	0.262	1.660	5.979	0.435	1.565
23	0.626	0.162	0.633	0.989	1.011	0.545	1.455	0.539	1.438	3.858	0.716	0.259	1.711	6.006	0.443	1.557
24	0.612	0.157	0.619	0.989	1.011	0.555	1.445	0.549	1.429	3.895	0.712	0.257	1.759	6.032	0.452	1.548
25	0.600	0.153	0.606	0.990	1.010	0.565	1.435	0.559	1.420	3.931	0.708	0.254	1.805	6.056	0.459	1.541

4. VALORES DE LOS RANGOS DE LOS EJEMPLOS DE LIU

```
Liu.ranks<-c(0.082, 0.948, 0.84, 0.256, 0.67, 0.074, 0.196, 0.392, 0.358, 0.296,  
0.022, 0.888,0.812, 0.852, 0.28, 0.446, 0.658, 0.068, 0.962, 0.348, 0.502, 0.472,  
0.186, 0.542, 0.566, 0.114, 0.424, 0.022, 0.102, 0.576, 0.866, 0.8, 0.768, 0.528,  
0.42, 0.444, 0.656, 0.436, 0.92, 0.876, 0.022, 0.022, 0.022, 0.194, 0.022, 0.1,  
0.022, 0.022, 0.194, 0.29, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022,  
0.022, 0.022, 0.588, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.202, 0.022, 0.896,  
0.28, 0.16, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022)
```